THREE ESSAYS ON COMMODITY MARKETS AND HEALTH ECONOMICS

A Dissertation

by

SIHONG CHEN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Ximing Wu
Co-Chair of Committee,	Yu Zhang
Committee Members,	David Bessler
	Qi Li
Head of Department,	C. Parr Rosson III

May 2017

Major Subject: Agricultural Economics

ABSTRACT

This dissertation includes three essays. The first essay examines time-varying nonlinear dependence and asymmetries of commodity futures from 1999 to 2015. We consider several elliptical copulas with dynamic conditional correlation (DCC) and block dynamic equicorrelation (Block DECO) to capture dependence structure of various commodities across different sectors. Our major findings include: (1) flexible copula specification that allows for multivariate asymmetry and tail dependence appears to have the best model performance in characterizing co-movements of commodity returns. (2) dynamic correlations reveal connectedness degree between commodities has dramatically increased during the financial distress and the European debt crisis, but they declined sharply after 2012 and returned to the pre-crisis level since. (3) conditional diversification benefit is disappearing and lower tail dependence between commodity markets is much higher in the bearish market.

The second essay studies volatility spillover and various connectedness measures for 20 commodity futures from 1996 to 2016. We propose to estimate network connectedness in commodity markets by a previous framework that models direction and magnitude of volatility spillover using reduced-form vector autoregression (VAR) models and generalized forecast error variance decomposition. We find clustering of commodity futures that match their industrial groupings, and energy markets have played a central role in the network in the static analysis. Our dynamic models show that though market interconnections have dramatically increased during the 2007-2009 financial crisis, they have returned to the pre-crisis levels after. We also find that recent downward movement of crude oil prices does not necessarily lead to stronger connectedness between commodity markets.

The third essay investigates health economics in developing countries. Obesity and overweight problems have become prevalent in developing countries like China. This paper presents a comprehensive analysis on body mass index (BMI) using a micro-level data of Chinese families. We model the dynamics of BMI determinants spanning from 1991 to 2011 for rural and urban residents. Our identification strategies include: (1) using spousal and parental characteristics as proxy variables to control for omitted variables bias and (2) explicitly modeling common couple effect with correlated random-effects regressions for spousal BMI. Our results find strong and positive spousal/intergenerational transmissions of BMI for families across region and time. Depending on the gender of spouse and grown children, besides transmission effects a variety of socioeconomic variables are identified as significant predictors of individual BMI.

DEDICATION

To my mother, my father, and my grandparents.

ACKNOWLEDGEMENTS

First and foremost, I want to thank my committee chair, Dr. Ximing Wu and co-chair Dr. Yu Zhang. They made it possible for me to write this dissertation and stood by my side while I began my research in applied economics. I want to thank Dr. Wu and Dr. Zhang for all the time we spent in their office discussing all the questions I had, correcting my mistakes, making improvements and sharing their ideas with me. I am very grateful to them for all the opportunities and financial support they gave me during the past five years.

I also want to express my gratitude to Dr. David Bessler and Dr. Qi Li for their continuous support and enlightenment. I thank them for reading previous drafts of this dissertation and providing valuable comments that improved the contents of this dissertation. Last but not least, I thank my family for their support during my Ph.D. study. This dissertation would not have been completed without their encouragement.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professor Ximing Wu, Professor Yu Zhang and Professor David Bessler of the Department of Agricultural Economics and Professor Qi Li of the Department of Economics.

The data analyzed for Chapter 4 was provided by Professor Shaosheng Jin of the Department of Agricultural Economics at Zhejiang University.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a departmental scholarship from the Department of Agricultural Economics at Texas A&M University.

NOMENCLATURE

CDF	cumulative distribution function
DCC	dynamic conditional correlation
Block DECO	block dynamic equicorrelation
BMI	body mass index
OOS	out-of-sample
CDB	conditional diversification benefit
CRE	correlated random effect
VAR	vector autoregression
ARIMA	autoregressive integrated moving average model
GARCH	generalized autoregressive conditional heteroskedasticity
ETF	exchange traded funds

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

Commodity futures have become an increasingly important component in the financial markets. Financialization of commodity markets, which began since 2004, has provided another option for managers of hedge fund and mutual fund to construct a more diversified portfolio. However, few previous research has paid attention to the time-varying dependence structure and co-movements of commodity markets, which is of great concern from the perspective of risk management. How do commodity markets move together during the financial crisis? Which commodities are major players in the markets? What are the impacts of sharp decline in crude oil price since 2014? To the best of my knowledge, these interesting problems remain unknown in the literature. I attempt to provide some insights for these problems using dynamic copulas and network connectedness models. Copulas is a powerful and flexible tool in modeling dependence structure of multivariate data. In the first essay, I employ a recent technique of dynamic copulas to investigate how dependence structure of commodity markets is changing over time. In the second essay, I adopt a vector autoregression (VAR) based approach to visualize network connectedness in the commodity markets. Both essays suggest consistent and robust findings though the models and data under consideration are completely different.

Obesity in developing country is another important but less explored topic in the literature. In my third essay I examine how body mass index (BMI), a useful measure for classification of obesity and overweight, is determined by socioeconomic and demographic variables in Chinese families. Particularly, I model the effects of spousal and intergenerational transmission of BMI on individual BMI over a span of two decades in China using reduced-form and structural regressions. Empirical results imply the effects of various variables on BMI are depending on gender, time and region.

In the rest of my dissertation, the first two essays will focus on co-movements and integration of commodity markets, and the last essay will specialize in empirical determinants of BMI in Chinese families across time and region. Some technical details of model and estimation are included in the appendix.

2. MODELLING DEPENDENCE STRUCTURE OF COMMODITY MARKETS WITH DYNAMIC COPULAS

2.1 Introduction

Commodity markets have attracted much attention in both academia and industry since early 2000s, but some interesting problems still remain unknown and are little studied so far. First, commodity exchange traded funds (ETF), which track specific commodity indices and invest in several different commodities, are rising as the sole focus of many institutional investors' portfolios since 2000s. Subsequent large inflows into commodity markets, termed financialization of commodity markets, are claimed to have substantially increased correlations between a large number of commodity futures (Tang and Xiong 2012; Büyükşahin and Robe 2014). Though recent research provides some evidence of structural change in correlation, most of them do not fully account for the information of higher moments or model the joint distribution of futures returns, partly due to the paucity of flexible multivariate distributions in the literature.

Second, many assets like stocks, bonds and commodities that showed low correlations in the history are often used to build well-diversified portfolios in mutual funds, but these assets have shown tendence to crash together during the 2008-2009 financial crisis. Therefore, understanding time-varying co-movements of a large collection of commodities futures is of great importance to construct an optimized dynamic portfolios from the perspective of risk management (Belousova and Dorfleitner 2012; Bessler and Wolff 2015; Daskalaki and Skiadopoulos 2011). However, few previous studies have provided a comprehensive analysis on diversification benefits of commodity futures markets over time.

There is a fast growing body of recent literature on interconnection of commodity markets or the role of financialization in markets co-movements. Saghaian (2010) presents empirical results using vector autoregression (VAR) and Granger causality supplemented by a directed graph theory modeling approach to identify the links and plausible contemporaneous causal structures between energy and commodities in the grain sector (wheat, soybean and corn). Although Saghaian (2010) finds strong correlation among oil and food prices with monthly data from 1996 to 2008, there is mixed evidence of a causal link from oil to the other three commodities. Büyükşahin and Robe (2017) model dynamic correlations between equity market and commodity in grains and livestock sector, and find that world business cycle shocks have a substantial and long-lasting impact on the food markets co-movements with equity market, while changes in the intensity of financial speculation have a short-lived and not significant impact on cross-market return linkages using various specifications of structural vector autoregression (SVAR). Tang and Xiong (2012) find increasing correlation since 2004, but they model dynamics of correlations by rolling-window for all pairwise combinations of commodities one after another, which is inefficient as they do not explicitly take all information into account and not necessarily robust to the structural change in correlations. Adams and Glück (2015) consider structural breaks in correlations but their sample only include eight commodities and also do not provide a joint estimation of dependence structure in futures returns. Most of these studies, however, only focus on specific commodities or just use low frequency data (monthly or weekly), and one may want to know if these empirical results are still robust if relative high frequency information of more futures markets is used.

The aim of this article is to explore whether the dependence structure of commodity markets is asymmetric and changing over time. To this end, we make use of copula, an extremely useful and informative approach to model multivariate distribution with a set of returns constructed from any flexible specification of marginal distributions. The last two decades have witnessed numerous copulas applications in modeling joint distribution of default probability of credit products in the finance industry, but few of them can successfully model dynamics of joint distribution and high dimensional data at the same time (Patton 2013). In recent years a few studies have proposed some new families of copulas to capture: (1) co-movements of a large number of equity returns and, (2) dynamic dependence structure that is robust to various financial and economic circles. For example, Christoffersen et al. (2012) propose a new class of dynamic copulas based on the dynamic conditional correlation of Engle (2002) and the multivariate skewed t distribution of Demarta and McNeil (2005) to model co-movements of asset returns in developed and developing markets. Christoffersen et al. (2012) find that diversification benefits from a large portfolio of these markets returns have gradually decline since 1970s. Creal and Tsay (2015) introduce time-variation into the copula densities by writing them as factor models with stochastic loadings. The proposed copula models have flexible dynamics and heavy tails yet remain tractable in high dimensions due to their factor structure. Creal and Tsay (2015) use Bayesian estimation approach that leverages a recent advance in sequential Monte Carlo methods known as particle Gibbs sampling to draw large blocks of latent variables efficiently and in parallel. They find strong evidence of time-varying correlation and tail dependence from a 200-dimensional unbalance panel of credit default swaps (CDS) and equity returns. Lucas, Schwaab, and Zhang (2017) develop a modeling framework of copulas based on the multivariate skewed t distribution of Demarta and McNeil (2005) and score-driven dynamic of Creal, Koopman, and Lucas (2013) to estimate joint and conditional tail risk probabilities over time in a financial system that consists of a large number of financial sector firms. Oh and Patton (2016) combine the generalized autoregressive score model

(GAS) of Creal, Koopman, and Lucas (2013) and the factor copula model of Oh and Patton (2017) to obtain a tractable and parsimonious time-varying model for high-dimensional conditional distributions. Oh and Patton (2016) use the proposed copulas to study 100 daily CDS spreads on the U.S. companies over the period 2006 to 2012, and find that while the probability of distress for individual company has significantly reduced since the financial crisis, the level of systemic risk is substantially higher than in the pre-crisis period. Amongst all existing copulas research in finance, to the best of our knowledge few has considered the joint dynamics of commodity markets in various sectors.

In this study we make three primary contributions to the current literature. First, we explore time-varying dependence structure of 23 commodities using dynamic copulas of Christoffersen et al. (2012) and show strong evidence of multivariate asymmetry and tail behavior in commodity returns. Second, our robust estimates of dynamic copula correlations reveal that commodity markets are most connected during the financial crisis, but these correlations have returned to the pre-crisis level after 2012. Particularly, idiosyncratic commodity shocks like those from food or energy markets do not necessarily imply higher correlations between commodity futures. Third, we find diversification benefit is disappearing and the tail dependence is substantially higher in the bearish market, but the optimal portfolio weights based on dynamic copulas always have better performance compared to the equal-weighted portfolio.

The paper proceeds as follows. Section 2.2 introduces the marginal model and skwed t copulas with two dynamic correlations. Section 2.3 discusses how we construct rolling future contracts to meet data requirement for dynamic copulas, and presents major empirical analysis on how copula correlations are changing over time. Out-of-sample model comparison is undertaken to shed light on predictive performance of various dynamic models. Section 2.4 presents two important economic implications of estimated dynamic copulas, including diversification benefit and lower tail dependence. Section 4.4 concludes. Introduction of multivariate skewed t distribution and implementation details of out-of-sample based model comparison are included in Appendices A and B.

2.2 Dynamic copula models

In this section we present the basics of our modeling framework for dynamic copulas in three parts. Section 2.2.1 describes how we estimate the marginal distribution of futures returns. Section 2.2.2 introduces skewed-t copula model that is widely used in finance literature in recent years. Section 2.2.3 presents the dynamic conditional correlation (DCC) model that describes time-varying dependence structure of futures returns. We briefly discuss how elliptical copulas and various specifications of dynamic correlations can be integrated using maximum composite likelihood estimation for data in high dimensions.

2.2.1 Marginal model

We assume $y_{i,t}$ is the log returns of commodity future i at period t, dynamic mean $\mu_{i,t}$ is captured by autoregressive integrated moving average (ARIMA) model, volatility $\sigma_{i,t}$ follows generalized autoregressive conditional heteroskedasticity (GARCH) process, and $\epsilon_{i,t}$ is the innovation term. The order of ARIMA model is selected by the Bayesian Information Criterion (BIC), and the GARCH(1,1) model is estimated by QMLE. The univariate model is described as:

$$(2.1)\quad y_{i,t} = \mu_{i,t} + \sigma_{i,t}\epsilon_{i,t}, \quad i = 1, ..., n,$$

$$(2.2)\quad \sigma_{i,t}^2 = \omega_i + \alpha_i(y_{i,t} - \mu_{i,t})^2 + \beta_i\sigma_{i,t-1}^2$$

(2.3) $\epsilon_{i,t} \sim Skew \;\; t(v_i, \lambda_{c,i})$

and the skewed-t distribution is defined in Hansen (1994):

$$(2.4) \;\; g(\epsilon|v,\lambda_c) = \begin{cases} bc\left(1 + \frac{1}{v-2}(\frac{b\epsilon+a}{1-\lambda_c})^2\right)^{-(v+1)/2}, & x < -a/b, \\ bc\left(1 + \frac{1}{v-2}(\frac{b\epsilon+a}{1+\lambda_c})^2\right)^{-(v+1)/2}, & x \geq -a/b. \end{cases}$$

and the constants a, b and c are given by

$$(2.5) \;\; a = 4\lambda_c c\left(\frac{v-2}{v-1}\right), \;\; b^2 = 1 + 3\lambda_c^2 - a^2, \;\; c = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{\pi(v-2)\Gamma(\frac{v}{2})}}$$

where $\lambda_c \in [-1, 1]$ is a skewness parameter which controls degree of asymmetry and $v \in (2, \infty]$ is degree of freedom which controls the thickness of tails. This skewed-t distribution is flexible and able to nest many special cases for marginal distribution. For example, $\lambda_c = 0$ implies it is a standardized student's t distribution, $v \to \infty$ suggests it is a skewed normal distribution, and if $v \to \infty$ and $\lambda_c = 0$ it becomes a normal distribution $N(0, 1)$. In our empirical analysis we estimate the marginal distribution using this ARIMA-GARCH model for each commodity futures returns. We obtain the cumulative distribution function (CDF) by

(2.6) $\eta_{i,t} \equiv \eta_{i,t}(\epsilon_{i,t}|\lambda_{c,i}, v_i)$

where $\lambda_{c,i}$ and v_i are the estimated parameters for the distribution of innovation terms $\epsilon_i = [\epsilon_{i,1}, ..., \epsilon_{i,T}]'$, which remain unchanged over time.

2.2.2 Skewed t *copula*

Elliptical copula models, say, normal and student's t copulas, are the most popular approach for modeling dependence structure of high dimensional data in finance for interpretation simplicity and convenient implementation. However, elliptical copulas also have some drawbacks that restrict their flexibility to model comovements in financial markets. For example, normal copula, which is derived from the multivariate normal distribution, does not assume any tail behavior in equity returns. Although t copula allows for tail dependence, it assumes the dependence is symmetric on both tails. During the 2007-2009 financial distress, however, we observed the equity returns tend to crash together, not to boom together, which apparently calls for a more flexible model to allow for asymmetric tail dependence in the multivariate data. Archimedean copula is another popular copula family that includes the Frank, the Clayton, the Gumbel and many other copula specifications, but this family is most successful in modeling bivariate data and difficult to generalize to high dimensional data. We refer interested readers to Patton (2013) for a comprehensive introduction of various copula models.

In this paper we follow Christoffersen et al. (2012); Christoffersen, Lunde, and Olesen (2014); Christoffersen et al. (2016) who consider the skewed t copula approach proposed in Demarta and McNeil (2005) to model high dimensional dependence in equity markets. To be specific, the skewed t copula is defines as:

$$C(\eta_{i,1}, ..., \eta_{i,N}; \Gamma, \lambda, v_c) = t_{\Gamma,\lambda,v_c}(t^{-1}_{\lambda_1,v_c}(\eta_{i,1}), ..., t^{-1}_{\lambda_N,v_c}(\eta_{i,N})) \tag{2.7}$$

where t_{Γ,λ,v_c} is the multivariate skewed t distribution, skewing parameter is an N-dimensional vector $\lambda = (\lambda_1, ..., \lambda_N)'$, v_c as degree of freedom, and Γ as correlation matrix. $t^{-1}_{\lambda_i,v_c}$ is the quantile function for the univariate skewed t distribution. Skewed t copula has a very flexible specification based on the multivariate skewed t distribution and nests t copula (when $\lambda \to 0$) and normal copula (when $\lambda \to 0$ and $v_c \to \infty$). Demarta and McNeil (2005) define the probability density function (pdf) of the N dimensional skewed t copula from the asymmetric t distribution as:

$$c(\eta|\Gamma, \lambda, v_c) = \frac{2^{\frac{(v_c-2)(N-1)}{2}} K_{\frac{v_c+N}{2}}(\sqrt{(v_c + \epsilon^{*\prime}\Gamma^{-1}\epsilon^*)\lambda'\Gamma^{-1}\lambda})e^{\epsilon^{*\prime}\Gamma^{-1}\lambda}}{\Gamma\left(\frac{v_c}{2}\right)^{1-N} |\Gamma|^{\frac{1}{2}} \left(\sqrt{(v_c + \epsilon^{*\prime}\Gamma^{-1}\epsilon^*)\lambda'\Gamma^{-1}\lambda}\right)^{-\frac{v_c+N}{2}} \left(1 + \frac{1}{v_c}\epsilon^{*\prime}\Gamma^{-1}z\epsilon^*\right)^{\frac{v_c+N}{2}}} \tag{2.8}$$

$$\times \prod_{i=1}^{N} \frac{\left(\sqrt{(v_c + \epsilon_i^{*2})\lambda^2}\right)^{-\frac{v_c+1}{2}} \left(1 + \frac{\epsilon_i^{*2}}{v_c}\right)^{\frac{v_c+1}{2}}}{K_{\frac{v_c+1}{2}}\left(\sqrt{(v_c + \epsilon_i^{*2})\lambda_i^2}\right) e^{\epsilon_i \lambda_i}}$$

where η is an N dimensional vector of cumulative distribution function, $K_{\frac{v_c+d}{2}}$ is the modified Bessel function of the third kind, and ϵ^* is given by

(2.9) $\epsilon^* = [t^{-1}_{\lambda,v_c}(\eta_1), ..., t^{-1}_{\lambda,v_c}(\eta_N)]$

Notice that ϵ^* is obtained via the quantile function of skewed t distribution, and ϵ is obtained directly from the ARIMA-GARCH estimation for univariate data. As a result, if the marginal distribution F_i is close to the univariate skewed t distribution of Demarta and McNeil (2005), ϵ^* is close to ϵ as well. As we will see shortly in next section, we need ϵ^* since it drives estimation of dynamic conditional correlation in the copula models. We consider using skewed t copula to measure comovements in commodity markets as it is highly tractable and flexible in modeling dependence structure even for hundreds of equity returns (Christoffersen, Jacobs, Jin, and Langlois 2016). For the sake of simplicity in our empirical analysis we will assume λ as a scalar for the high-dimensional commodity returns. Notice that λ is used to capture multivariate asymmetry of skewed t copula while $\lambda_{c,i}$ in the marginal model is used to capture skewness in the univariate returns.

2.2.3 Dynamic conditional correlation for copulas

Motivated by the seminal paper of Engle (2002) and a recent application of Christoffersen et al. (2014), we propose to integrate dynamic conditional correlation (DCC) process with skewed t copula to capture correlation dynamics, asymmetry, and tail behavior in the multivariate distribution of commodity returns. Since an original DCC process is driven by a multivariate GARCH process and the copula shocks $\epsilon^* = [t^{-1}_{\lambda,v_c}(\eta_1), ..., t^{-1}_{\lambda,v_c}(\eta_N)]$ used to drive dynamic correlation do not necessarily have zero mean and unit variance from the skewed t copula model, we need to standardize these copula shocks before modeling correlation dynamics. Integrating DCC process with copulas implies standardization is still needed for t copulas*. Notice that standardized copulas shocks are only used in the DCC process, in the maximum composite likelihood estimation we still use the original copula shocks in the likelihood function†.

We assume the dynamic conditional correlation for elliptical copulas is driven by a GARCH-type process:

$$(2.10)\quad \tilde{\Gamma}_t = (1 - \alpha_\Gamma - \beta_\Gamma)\Omega + \alpha_\Gamma \bar{\epsilon}^*_{t-1} \bar{\epsilon}^{*'}_{t-1} + \beta_\Gamma \tilde{\Gamma}_{t-1}$$

where α_Γ and β_Γ are scalars, and $\bar{\epsilon}^*_t$ is an N dimensional vector at period t such that $\bar{\epsilon}^*_{i,t} = \epsilon^*_{i,t}\sqrt{\tilde{\Gamma}_{ii,t}}$. We use equation (2.10) to define the conditional correlation by the following normalization to ensure correlations are always in the $[-1, 1]$ interval:

$$(2.11)\quad \Gamma_{ij,t} = \tilde{\Gamma}_{ij,t} / \sqrt{\tilde{\Gamma}_{ii,t}\tilde{\Gamma}_{jj,t}}$$

We transform $\epsilon^*_{i,t}$ to $\bar{\epsilon}^*_{i,t}$ as Aielli (2013) points out this change ensures a consistent estimate of Ω for moment matching, which defines

$$(2.12)\quad \hat{\Omega} = \frac{1}{T}\sum_{t=1}^{T} \bar{\epsilon}^*_{i,t}\bar{\epsilon}^{*'}_{i,t}$$

since Ω is a copula correlation matrix, all diagonal elements of Ω equal one and we only need $\tilde{\Gamma}_{ii,t}$ for all i to yield $\bar{\epsilon}^*_{i,t}$. Aielli (2013) shows that we can first obtain the diagonal elements of equation (2.10) for all i and t by

$$(2.13)\quad \tilde{\Gamma}_{ii,t} = (1 - \alpha_\Gamma - \beta_\Gamma) + \alpha_\Gamma \bar{\epsilon}^{*2}_{t-1} + \beta_\Gamma \tilde{\Gamma}_{ii,t-1}$$

*In the appendix we show how to standardize copula shocks for t and skew t copulas. Standardization is not needed for normal copula since the shocks have zero mean and unit variance.

†The copula likelihood function is not affected by the DCC process.

which can be used to compute $\bar{\epsilon}^{*}_{i,t}$ to yield $\hat{\Omega}$ from equation (2.12). This is a direct modeling of correlation dynamics and has the potential to capture precisely the time-varying nature of correlations. In a recent paper Engle and Kelly (2012) extend dynamic conditional correlation to dynamic equicorrelation (DECO) and block dynamic equicorrelation (Block DECO) to further reduce computational burden and utilize potential group information in the data. As its name suggests, DECO implies equal correlation between any pairwise combination of data at period t, and Block DECO implies the data belong to different groups hence the correlation matrix has a group structure and time-varying elements at each period. Engle and Kelly (2012) show that consistent estimates of DECO and Block DECO are straightforward based on dynamic conditonal correlation Γ_t from equation (2.11):

$$(2.14)\quad \Gamma_{\text{DECO},t} = \left(\frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j>i} \Gamma_{ij,t} \right) \boldsymbol{I_{N \times N}}$$

and the off-diagonal ij^{th} element of Block DECO at period t is:

$$(2.15)\quad \Gamma_{\text{Block DECO},ij,t} = \begin{cases} \frac{2}{N_k(N_k-1)} \sum_{r,s \in k, r \neq s} \Gamma_{rs,t}, & i,j \in \text{group } k, \\ \frac{1}{N_k \times N_l} \sum_{r \in k, s \in l} \Gamma_{rs,t}, & i \in \text{group } k, \ \ j \in \text{group } l. \end{cases}$$

where N_k and N_l represent number of members in group k and group l respectively. We are particularly interested in Block DECO as an alternative specification for dynamic correlation in elliptical copulas as commodity futures have been traditionally clustered by their industrial groupings and show evident group behaviors across financial and economic circles, while as far as we are concerned there is few previous research that takes group behavior of futures into account when modeling their dynamic dependence structure.

Engle, Shephard, and Sheppard (2008) and Engle and Kelly (2012) propose to estimate DCC, DECO and Block DECO model for high dimensional applications with composite likelihood to reduce computational burden with full likelihood. Christoffersen et al. (2012, 2016) adopt this approach and demonstrate it is highly reliable to model dependence structure with up to hundreds of variables using dynamic copulas.

To be specific, the composite likelihood in our context is defined as:

$$CL(\lambda, v_c, \alpha, \beta) = \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j>i} \ln f(\lambda, v_c, \alpha, \beta; \epsilon_{i,t}^*, \epsilon_{j,t}^*) \tag{2.16}$$

where $f(\alpha, \beta; \epsilon_{i,t}^*, \epsilon_{j,t}^*)$ denotes a bivariate elliptical copula density of pair i and j with correlation defined from either DCC, DECO or Block DECO. We maximize composite likelihood by summing over all possible pairs in each period t, and find this consistent estimator numerically fast and efficient. In next section we combine normal, student and skewed t copulas with DCC and block DECO to capture dynamic dependence structure in the commodity returns[‡], which eventually yields six time-varying copulas for our empirical analysis.

2.3 Empirical results

In this section we present major results for dynamic copulas estimation and out-of-sample based model comparison. In Section 2.3.1 we introduce how we construct the rolling commodity prices to handle futures with various expiry dates. Section 2.3.2 discusses the estimation results of dynamic copulas with various specification of elliptical copula and dynamic correlation. In Section 2.3.3 we use prediction comparison test to rank performance of dynamic copulas based on out-of-sample copula density.

[‡]We have considered DECO for dynamic copulas but both of its in-sample and out-of-sample performance are significantly worse than DCC and block DECO, therefore we omit these results for the sake of parsimony.

2.3.1 Data

We consider 23 commodities in the Goldman Sachs Commodity Index (GSCI) that serves as a benchmark for investment in the commodity markets. In particular we have 3 commodities in energy sector (WTI crude oil, Brent crude oil and natural gas), 6 commodities in grains sector (corn, soybean, soybean oils, oats, wheat and rough rice), 6 commodities in softs sector (coffee, cotton, sugar, cocoa, lumber and orange juice), 3 commodities in livestocks sector (feeder cattle, lean hogs and live cattle) and 5 commodities in metals sector (gold, platinum, palladium, silver and copper). We use these five sectors to determine the grouping in Block DECO copulas.

We collect the dataset from Datastream using ticker "CS04" from 29th Dec 1998 to 23rd May 2015 for all commodity futures. This ticker stands for continuous returns which are rolled when the first-nearest to expire future contracts have reached expiry date. On this date the second-nearest to expire future contracts returns are used to ensure these futures returns are all based on the same contract. We focus on this period as the commodity markets have experienced so called "financialization" after 2004 and gone through different financial and economic circles since 2008, therefore understanding how dependence structure of markets is varying over time is a key to diversify systemic risk of a large portfolio with commodity futures.

2.3.2 Estimation results of dynamic copula models

From Section 2.2 we know that estimation of dynamic copulas is actually a two-stage process. In the first stage we use quasi maximum likelihood estimation (QMLE) to obtain GARCH volatility of commodity returns, and in the second stage we use estimated marginal distribution and the proposed maximum composite likelihood estimation to obtain α, β, λ and v_c that drive the dynamics of high dimensional time-varying copulas. For the sake of parsimony we omit the univariate GARCH results for all 23 commodities below and focus on DCC and Blcok DECO copulas. Table 2.1 below reports parameter estimates for normal, student and skewed-t copulas:

Table 2.1: Parameter estimates for dynamic copula models

	α	β	Persistence	v_c	λ	Likelihood
DCC Copulas						
Normal	0.00684**	0.99023**	0.99707	-	-	28628.29
	(0.00018)	(0.00029)				
Student	0.00682**	0.99044**	0.99726	29.420**	-	29287.55
	(0.00018)	(0.00030)		(0.32457)		
Skewed t	0.00681**	0.99041**	0.99722	29.208**	-0.16908**	29357.64
	(0.00013)	(0.00022)		(0.34000)	(0.00948)	
Block-DECO Copulas						
Normal	0.01319**	0.98333**	0.99652	-	-	22636.49
	(0.00052)	(0.00074)				
Student	0.01313**	0.98380**	0.99693	25.431**	-	23470.08
	(0.00061)	(0.00076)		(0.18905)		
Skewed t	0.01305**	0.98382**	0.99687	25.316**	-0.13518**	23527.83
	(0.00049)	(0.00072)		(0.49173)	(0.00338)	

Notes: This table reports parameter estimates for DCC based copula models and Block-DECO based copula models with full sample. Standard errors are reported below the estimated parameters. * and ** denote significance at the 5% and 1% level respectively.

where dependence persistence $(\alpha + \beta)$ represents degree of mean-reversion in copula correlations, and the last column of Table 2.1 stands for copula likelihood[§]. Table 2.1 shows that estimated parameters of all dynamic copulas are significant. Skewed t copula has a negative asymmetry parameter λ in both correlation specification, implying evidence of multivariate asymmetry, which is typically seen in stock returns. As v_c of student and skewed t copulas with same dynamic correlation are rather close, and the only difference between them is the existence of λ in skewed t copula, we

[§]The full likelihood function of copula model is computed with both copula likelihood and marginal likelihood for univariate returns, but since marginal models are identical across all six models we only report copula likelihood here.

think this is a strong signal to reject multivariate symmetry in commodity futures returns. Another interesting result in Table 2.1 is the dependence persistence being close to one in all models, which suggests mean-reversion in copula correlations is slow in equation (2.10). To better understand the empirical results we consider an interesting comparison between the estimates of Christoffersen et al. (2012) and ours. In Christoffersen et al. (2012) the estimated degree of freedom parameters v_c and asymmetry parameter λ from skewed t copulas for 16 developed markets, 13 emerging markets and all markets are $[17.64, 22.37, 21.83]$ and $[-0.48, -0.49, -0.41]$, respectively. Our estimates of both parameters are around 29 and -0.17, which suggest equity markets have fatter tails and more asymmetry than the commodity markets.

A simple but informative in-sample model selection procedure can be undertaken by comparing the composite likelihood of dynamic copulas. Copula likelihoods in Tabel 2.1 implies two important findings: (1) with identical correlation specification (DCC or Block DECO) skewed t copula is marginally better than student copula, and normal copula has the worst performance among all three elliptical copulas and, (2) DCC is preferred to Block DECO for all copula specifications. The first result should not be too surprising as skewed t copula is more flexible and able to capture both multivariate asymmetry and tail behavior of commodity returns. It is interesting to see the second result which suggests DCC copula generally has better performance than Block DECO copula, and we attribute this finding to the fact that DCC is an unrestricted specification for correlation while Block DECO is somewhat restricted by the number of groups. Although modeling dependence structure of 23 commodity futures is a high dimensional application in copulas literature, this dimension is still not high enough to fully recognize the benefits of Block DECO that employs grouping information[¶].

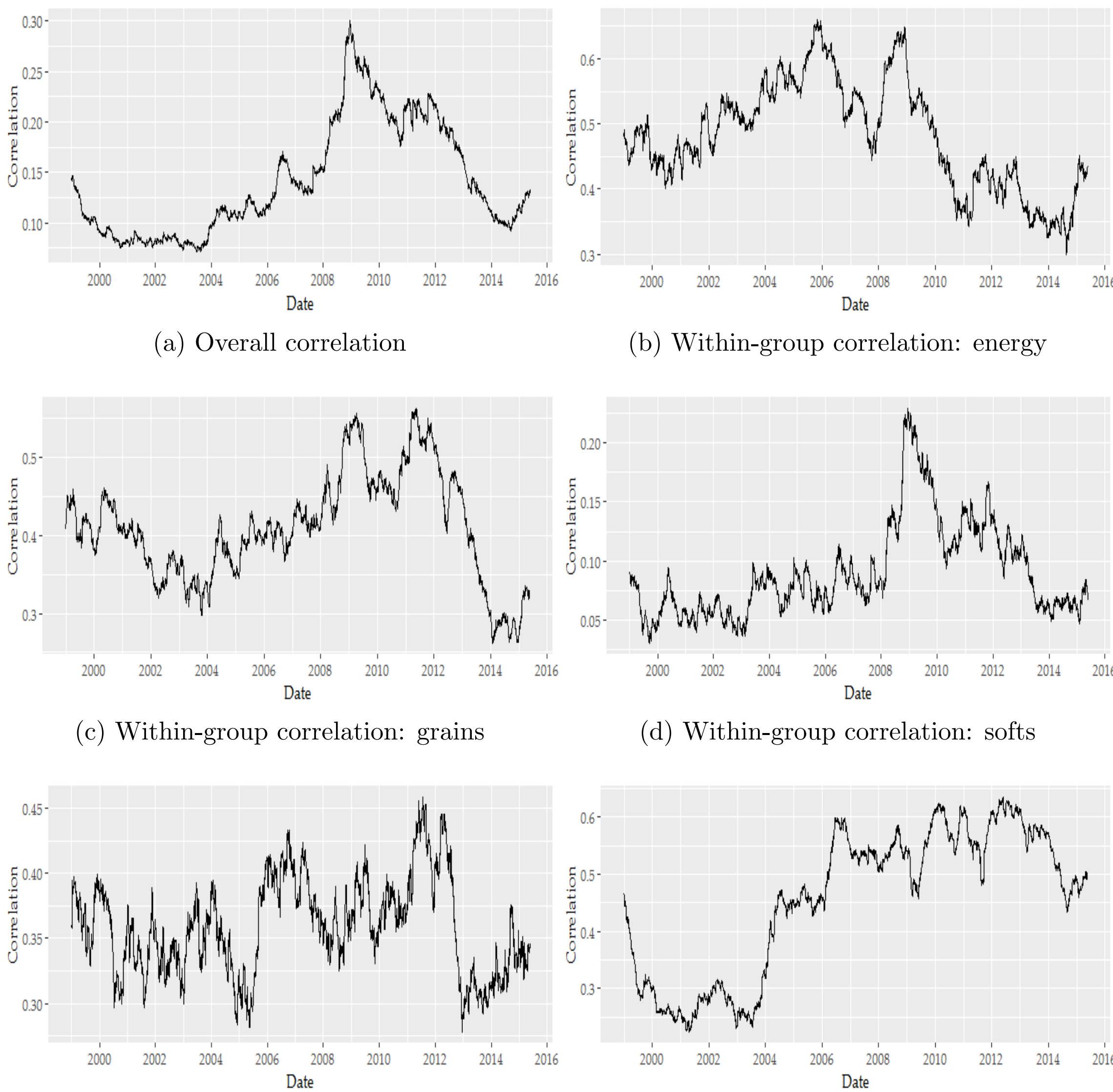
To investigate how dependence structure is changing over time, we show the estimated dynamic correlation from DCC skewed t copula in Figure 2.1, 2.2 and 2.3. Since there are $N(N-1)/2 = 253$ correlations at each period, it is extremely difficult to find any patterns with all correlations therefore we cluster commodities by their groups and present them at group level, reducing the number of within-group and cross-group correlations to 15 only. We take the average of all dynamic correlations at the same period t to have an overall dynamic dependence measure for the commodities under consideration.

Looking at Figure 2.1, 2.2 and 2.3, it is evident that most of the group-based correlations have increased and peaked during the 2007-2009 financial crisis, but all of them except metals group own correlation have declined sharply after the distress period. Most of them have returned to the pre-crisis level, which poses challenge on previous studies that claim commodity financialization since 2004 have increased dependence among a variety of commodity futures (Adams and Glück 2015; Tang and Xiong 2012). Since mid-2014 commodity markets have witnessed dramatic decreases in energy prices, and we find that overall correlation and cross-group correlations between energy and other groups have only risen very mildly. We suspect this is because decline in energy prices is mainly due to the positive supply shock in energy industry since 2014 rather than a negative demand shock that have widespread impacts on all markets (e.g., financial distress in 2007-2009).

2.3.3 Out-of-sample based model comparison

In this section we conduct model comparison based on out-of-sample (OOS) forecast of copula density. To test the predictive ability of dynamic copulas during the

[¶]Engle and Kelly (2012) show the power of Block DECO using all constituents of the S&P 500 Index



(a) Overall correlation

(b) Within-group correlation: energy

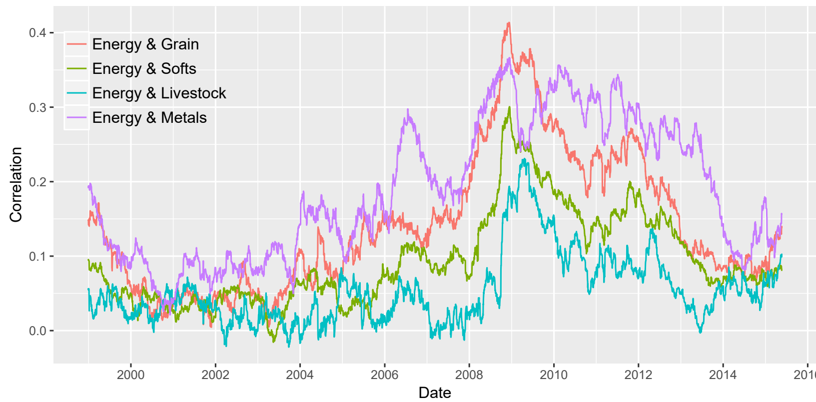
(c) Within-group correlation: grains

(d) Within-group correlation: softs

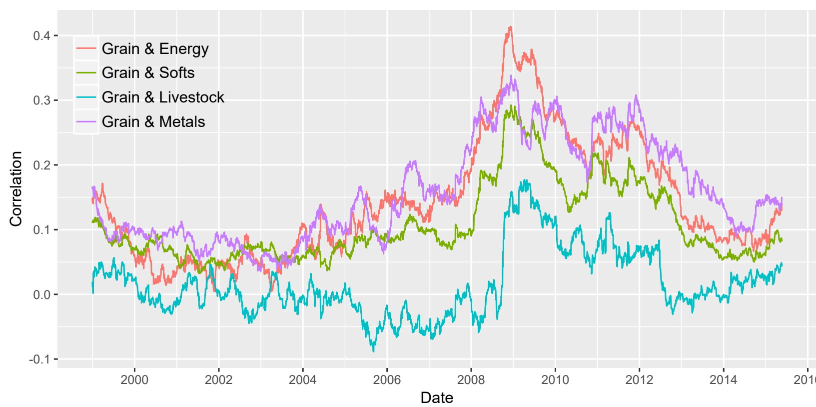
(e) Within-group correlation: livestocks

(f) Within-group correlation: metals

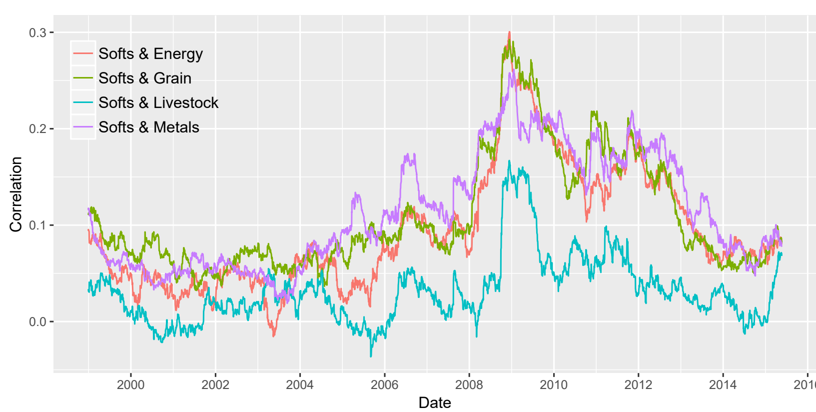
Figure 2.1: Dynamic correlations of skewed t copula: overall and within-group



(a) Group correlation of energy and other sectors

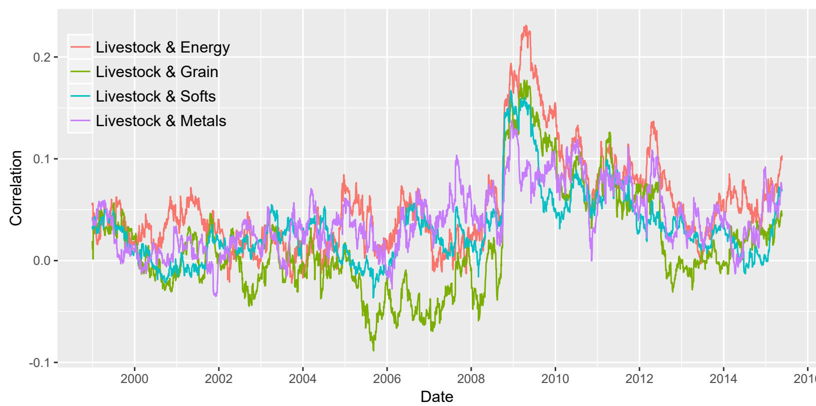


(b) Group correlation of grain and other sectors

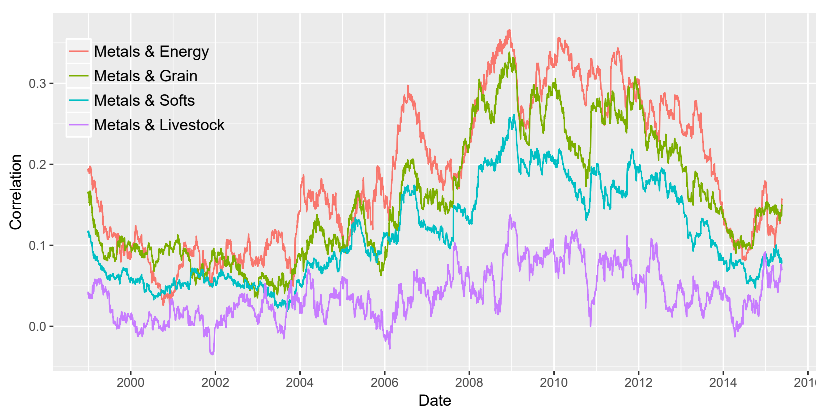


(c) Group correlation of softs and other sectors

Figure 2.2: Cross-group correlations of dynamic skewed t copula: energy, grain and softs sector



(a) Group correlation of livestock and other sectors



(b) Group correlation of metals and other sectors

Figure 2.3: Cross-group correlations of dynamic skewed t copula: livestock and metals sector

stress period and various economic circles in recent years, we use the data from June 1st 2007 to Dec 1st 2009 (the financial crisis period) as well as those from May 1st 2014 to May 23rd 2015 (sharp decline in crude oil prices since mid-2014), with a total of 900 days as the OOS test set and use a fixed rolling window of 1000 days to fit the dynamic copulas and predict one-day ahead copula density[‖]. We choose these days for OOS test as they represent volatile periods in the commodity markets, and one might want to know if the proposed dynamic copulas may capture market co-movements in time and prove useful with good prediction performances during these periods. Following Diebold and Mariano (1995), Giacomini and White (2006) and Patton (2013), we propose to compare two competing dynamic copulas conditional on the their estimated parameters:

$$
\begin{aligned}
H_0: &\quad E[\log c_1(\hat{\eta}_t;\hat{\theta}_{1t}) - \log c_2(\hat{\eta}_t;\hat{\theta}_{2t})] = 0 \\
\text{vs } H_1: &\quad E[\log c_1(\hat{\eta}_t;\hat{\theta}_{1t}) - \log c_2(\hat{\eta}_t;\hat{\theta}_{2t})] > 0 \\
H_2: &\quad E[\log c_1(\hat{\eta}_t;\hat{\theta}_{1t}) - \log c_2(\hat{\eta}_t;\hat{\theta}_{2t})] < 0
\end{aligned}
$$

where $\hat{\eta}_t$ is the predictive CDFs at period t, and $\hat{\theta}_{it}$ for $i = 1, 2$ is the estimated copula parameter from the fixed-length rolling window. The asymptotic framework of the predictive log-likelihood test is based on Giacomini and White (2006), which does not require any adjustments for the estimated parameters of the competing copulas and the limiting distribution of the test statistic is $N(0, 1)$. Giacomini and White (2006) show that as the differences in log-likelihoods is potentially heteroskedastic and serially correlated, we need to first estimate the heteroskedasticity and autocorrelation consistent (HAC) covariance for the predictive test. Diebold (2015) suggests this test statistic can be simply computed by regressing the differences in log-likelihoods on an intercept using HAC robust standard error. If t values of the intercept estimator are

[‖]See appendix for implementation details of the OOS based model comparison greater than 1.96 or less than -1.96, we may argue there is significant difference in predictive ability between competing models under consideration. We shows model comparisons results for all dynamic copulas in Table 2.2 below.

Table 2.2: Out-of-sample model comparisons for dynamic copulas

Correlation	DCC			Block DECO		
Copula	Normal	Student	Skewed t	Normal	Student	Skewed t
<u>DCC</u>						
Normal	-					
Student	3.47	-				
Skewed t	3.07	**1.08**	-			
<u>Block DECO</u>						
Normal	-14.39	-16.28	-15.45	-		
Student	-9.47	-14.55	-15.04	4.10	-	
Skewed t	-8.53	-13.16	-14.59	3.46	**0.67**	-
Model ranking	3	1	1	6	4	4

Notes: This table presents t-statsitics from out-of-sample pair-wise comparisons of the log-likelihood values for six dynamic copula models. A positive value suggests the model to the left is better than the model above, and a negative value suggests the opposite. t values which fall between $[-1.96, 1.96]$ are in bold.

In Table 2.2, 13 out of 15 model comparisons imply significant better performance of one dynamic copula against another one in terms of predictive copula density. The results are consistent with we have found in section 2.3.2, where all DCC copulas have much higher composite likelihood than the Block DECO copulas, and dynamic normal copula is beaten by student and skewed t copulas with identical correlation specification. We also confirm that skewed t copula is only marginally better than student copula as the t stats are 1.08 and 0.67 for DCC specification and Block DECO specification, respectively. Notice that the OOS model comparison test is only based on 900 observations during stress and volatile markets periods. We make this selection for OOS test on purpose as it is extremely difficult to forecast co-movements of commodity markets during these periods, and it would be interesting to see whether skewed t copula provides enough flexibility to model dependence structure and capture multivariate asymmetry. Although skewed t copula can better fit the data, we argue its predictive ability is not significantly better than student copula during volatile markets period.

2.4 Economic implications

This section presents two important economic applications of dynamic copulas. In section 2.4.1 we explore the benefits of portfolio diversification based on our estimate of DCC skewed t copula and a dynamic measure derived from expected shortfall by Christoffersen et al. (2012). In section 2.4.2 we measure lower tail dependence of dynamic copula to understand how systemic risk of commodity markets is varying over time.

2.4.1 Dynamic diversification benefit

The dynamic measure of diversification benefit is closely related to expected shortfall which is defined as:

$$(2.17)\quad ES_t^q(y_{i,t}) = -E[y_{i,t}|y_{i,t} \leq \eta_{i,t}^{-1}(q)]$$

where $\eta_{i,t}^{-1}(q)$ is the quantile function of returns i at period t, and q is a probability that we set as 5% in the following analysis. As Artzner et al. (1999) point out that expected shortfall has the sub-additivity property such that

$$(2.18)\quad ES_t^q(w_t) \leq \sum_{i=1}^{N} w_{i,t} ES_t^q(y_{i,t}), \ \text{ for all } \ w_t$$

where $ES_t^q(w_t)$ is the expected shortfall of a portfolio with weights w_t. As a result, the weighted average of returns' individual expected shortfalls is the upper bound on the portfolio's expected shortfall:

$$(2.19)\quad \overline{ES}_t^q(w_t) \equiv \sum_{i=1}^{N} w_{i,t} ES_t^q(y_{i,t})$$

which implies that a portfolio without any diversification benefits has a lower bound on expected shortfall:

$$(2.20)\quad \underline{ES}_t^q(w_t) \equiv -\eta_t^{-1}(w_t, q)$$

where $\eta_t^{-1}(w_t, q)$ is the quantile function for a portfolio with weight w_t. This lower bound represents an extreme case such that the portfolio return never being lower than its q^{th} distribution quantile. Christoffersen et al. (2012) propose a dynamic measure of diversification benefit as:

$$(2.21)\quad CDB_t(w_t, q) \equiv \frac{\overline{ES}_t^q(w_t) - ES_t^q(w_t)}{\overline{ES}_t^q(w_t) - \underline{ES}_t^q(w_t)}$$

which only takes values on the $[0, 1]$ interval and is not conditional on the level of returns $y_{i,t}$. Notice also that this measure is an increasing function in the degree of diversification benefit. In our analysis we maximize $CDB_t(w_t, q)$ by choosing optimal w_t with constraints $\sum_{i=1}^{N} w_{i,t} = 1$ and $w_{i,t} \geq 0$ for all i to mimic dynamic optimization process for portfolio rebalancing. Since the expected shortfall $ES_t^q(w_t)$ is not known in closed form based on dynamic copulas, therefore we follow Christoffersen et al. (2012) to measure $CDB_t(w_t, q)$ in three steps:

Step 1: simulate 5,000 returns for each commodity future at period t using the estimated univariate ARIMA-GARCH model and DCC skewed t copula with conditional returns, volatility and correlation that are available at period $t-1$.

Step 2: compute $\overline{ES}_t^q(w_t)$, $\underline{ES}_t^q(w_t)$ w_t and $ES_t^q(y_{i,t})$ for any given weights w_t using the simulated returns from step 1 and maximize $CDB_t(w_t, q)$ over w_t with constraints $\sum_{i=1}^{N} w_{i,t} = 1$ and $w_{i,t} \geq 0$ for all i.

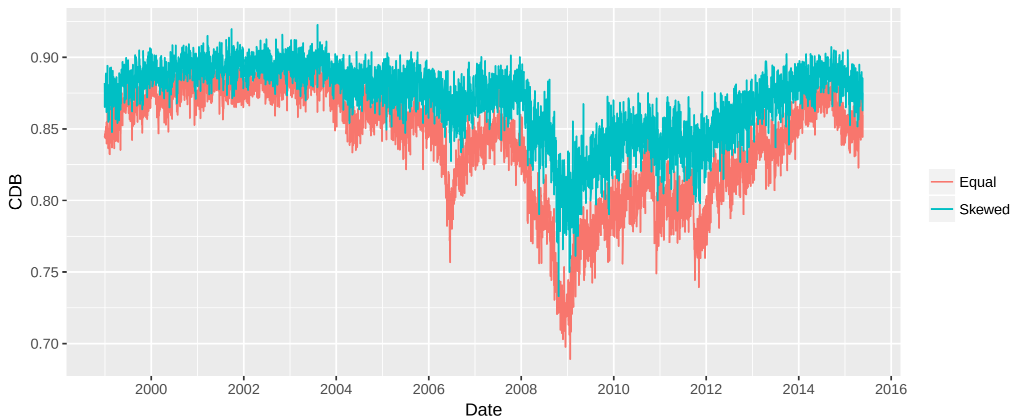


Figure 2.4: Conditional diversification benefits with skewed t copula

Step 3: save the optimal weights w_t and corresponding $CDB_t(w_t, q)$ and repeat the simulation and optimization process in step 1 and 2 for period $t+1$.

Figure 2.4 presents CDB results from our estimates in section 2.3.2. To compare the diversification benefits between DCC skewed t copula based portfolio and equal-weighted portfolio we denote the former one as "Skewed" and latter one as "Equal" in the following discussion. It is clear that CDB from both models are quite close before 2006. After 2006 CDB_t^{Equal} have declined sharply, though this pattern is also observed for CDB_t^{Skewed}, its decline is not as obvious as CDB_t^{Equal}. These decreases seem to be temporary and both benefit measures have gradually risen to the previous level before 2008. During the 2008-2009 financial crisis, diversification benefits have dropped markedly and reached the bottom at the beginning of 2009. Note also that the discrepancy between both CDBs are also greatest at this time. From 2009 CDB once again rose mildly and reached the pre-crisis level since 2014. From Figure 2.4 we argue that the value of skewed t copula is highest when commodity markets are volatile as it is able to capture multivariate asymmetry, nonlinear dependence and higher-order moments during the distress period when equity returns tend to crash together, not to boom together.

2.4.2 Dynamic tail dependence

One advantage of skewed t copula over other elliptical copulas is it allows for non-zero dependence and asymmetric dependence in the upper and lower tails, while normal copula assumes zero tail dependence and student copula implies symmetry on both tails. We are especially interested in the lower tail dependence as equity returns are usually much more connected and have higher co-movements in the bull market. Lower tail dependence is defined by the probability limit:

$$\tau^L_{i,j,t} = \lim_{\xi \to 0} \Pr[\eta_{i,t} \leq \xi | \eta_{j,t} \leq \xi] = \lim_{\xi \to 0} \frac{C_t(\xi, \xi)}{\xi} \tag{2.22}$$

where ξ is the tail probability. We note that the measure of lower tail dependence above is only bivariate and difficult to generalize to higher dimensions, and to display the dynamics of tail dependences among commodities futures we take the average of bivariate tail dependence across all pairs of commodities. As skewed t copula does not have an analytical solution for the lower tail dependence, we use numerical integration with $\xi = 0.001$ for approximation of copula function in equation (2.22).

Figure 2.5 shows the dynamic lower tail dependence averaged across all pairs of futures, which increased dramatically during the 2008-2009 financial crisis and reached the peak at the beginning of 2009. However, it declined rapidly from the peak but mildly rose again during the 2011-2012 European debt crisis. The lower tail dependence began to decrease after the debt crisis and return to the pre-criss level since 2014. Recall that this pattern is very consistent with what we have seen in section 2.4.1, where CDB_t went through a similar process during the crisis periods. Therefore, conditional diversification benefit and dynamic tail dependence appear to

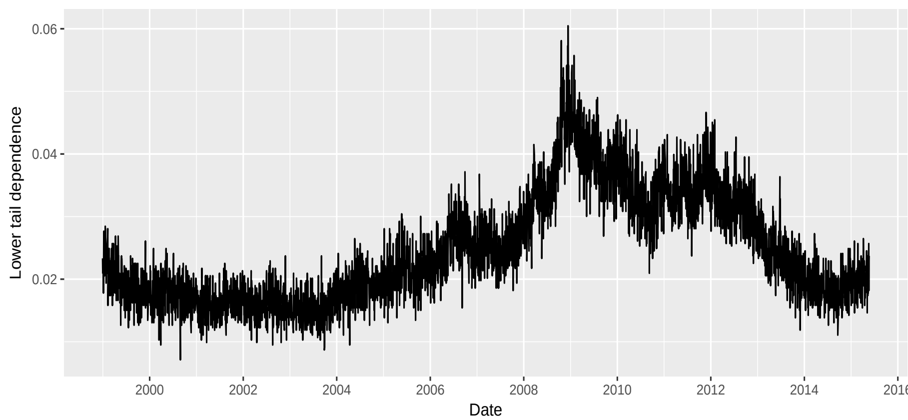


Figure 2.5: Dynamic average bivariate tail dependence for skewed t copula

be complementary measures for each other.

In Christoffersen et al. (2012) the dynamic measure of average lower tail dependence between 29 emerging and developed markets during 1989-2009 has trended upward, rising from almost 0 to around 0.18. Notice that 0.18 is much higher than our measure for commodity markets, which reached its peak of 0.06 during the financial crisis. This comparison suggests that though commodity markets have shown strong evidence of increasing dependence, its dependence level is much lower than the equity markets of various countries.

2.5 Concluding remarks

We model dynamic dependence structure of a large collection of daily commodity returns from various sectors over 1999-2015. Our robust copulas estimates reveal strong evidence of multivariate asymmetry, fat tails and time-varying co-movements in the commodity markets. We show that copula correlations between various commodities have increased substantially during the 2008-2009 financial crisis, but they returned to the pre-crisis levels after 2012. The dynamic copulas have two important economic implications. First, although conditional diversification benefits for a portfolio of commodity futures have declined dramatically during the crisis period, DCC skwed t copula based optimal portfolio has beaten equal-weighted portfolio by a large margin in the bearish market. Second, tail dependence of commodity futures, which is a useful measure for systemic risk, has trended upward in exactly the same period when diversification benefit reached the bottom. Our results suggest that dynamic copula is able to shed light on risk management to construct a well-diversified portfolio with a large number of commodity futures.

It may prove interesting to investigate and extend the models we use for other topics in the future. First, understanding how co-movements of commodity markets are affected by macroeconomy appears to be difficult, as macroeconomic variables are usually released at a relatively low frequency (monthly or quarterly) compared to asset returns. To handle this problem Ghysels, Santa-Clara, and Valkanov (2004) propose mixed-data sampling regression models (MIDAS) that are attracting growing interest in recent years. Combining dynamic copulas and MIDAS seem to be a promising approach to model interactions between macroeconomy and commodity markets without sacrificing much information in the data. Second, Diebold and Yilmaz (2014) develop a variety of connectedness measures of equity returns based on network topology and VAR models. One might want to know how and if the empirical results of dynamic copulas and connectednesss measures could be integrated to produce useful insights for risk management. To the best of our knowledge, there is no unifying framework to answer the problem. We conclude by raising these questions and hope they can be extended to a wider range of future studies.

3. NETWORK CONNECTEDNESS IN COMMODITY MARKETS

3.1 Introduction

Increasing connectedness among commodity markets have attracted much attention in both academia and industry since early 2000s, but some important problems remain unknown. First of all, although recent literature argues that large inflows into commodity markets since 2004, termed as "financialization of commodity markets", has dramatically raised interconnection between a large number of commodity futures (Tang and Xiong 2012), we are still far from consensus on the degree and evolution of commodity markets integration. As many assets like stocks, bonds and commodities that exhibit low correlation in the history tend to crash together during the recent financial crisis, understanding time-varying interdependence of a large number of commodity markets is of great importance to understand systematic risk in commodity futures. See Glasserman and Young (2016) for a comprehensive review of recent studies on contagion in financial network.

Second, most previous research concentrate on spillover effects of commodity returns, while transmission mechanism of volatility between commodities is still not clear. In a recent study Adams and Glück (2015) show significant risk spillover from stocks to commodities, but they only measure spillover using value-at-risk and do not consider how shocks in one market may affect volatility in another. By modeling volatility spillovers we may quantify the magnitude and direction of volatility shocks from various commodities with a causal interpretation, which is not readily available by modeling co-movements alone. As a result, measures of markets connectedness based on volatility spillovers of commodity futures may shed light on portfolio construction and diversification in real world practice.

A considerable body of literature has been devoted to examining spillover effects of volatility in financial markets. Diebold and Yilmaz (2012, 2014, 2015b) introduce and develop several connectedness measures from forecast error variance decomposition based on vector autoregression (VAR) models. They treat off-diagonal entries in variance decomposition as the influence of shocks to one market on the forecast error in another markets, and use these influences to build directed connection network of financial markets. They apply this framework to study volatility spillovers among U.S. and Global banks during financial crisis, and find the connectedness measures as useful monitor to precisely detect net volatility senders and receivers in the financial network. See Diebold and Yilmaz (2015a) for a book-length treatment on this new model.

Our research is most relevant to two recent papers that focus on volatility spillover between crude oil and agricultural products and evaluate the impacts of food crisis in 2006 on markets interconnection. Du, Yu, and Hayes (2011) investigate volatility spillovers of futures prices between crude oil and corn/wheat from November 1998 to January 2009. They use Bayesian Markov Chain Monte Carlo methods to estimate bivariate stochastic volatility based on the work of Duffie, Pan, and Singleton (2000) and Eraker, Johannes, and Polson (2003). Du, Yu, and Hayes (2011) show evidence of volatility spillover between crude oil and the other two agricultural products after the fall of 2006, and they attribute these findings to the results of ethanol production and increasing presence of index investors. Nazlioglu, Erdem, and Soytas (2013) examine price volatility transmission between oil and several agricultural commodity (wheat, corn, soybean and sugar) before and after the food price crisis in 2006. They apply the causality in variance test developed by Hafner and Herwartz (2006) to the GARCH volatility of commodity futures in pre-crisis and post-crisis periods separately. Nazlioglu, Erdem, and Soytas (2013) find significant volatility spillovers from crude oil markets to all agricultural markets only during post-crisis period, but they find wheat has significant spillover effect on oil in both periods. Although both studies show evidence of increasing volatility spillover from crude oil to agricultural commodities, they only consider a small number of products and do not take into account the recent decline of crude oil prices since 2014.

This paper attempts to identify direction and quantify magnitude of spillover effects in price volatility of 20 commodity futures covering energy, grains, softs, livestock and metals sectors. We make two primary contributions to the literature. First, we adopt the novel approach of Diebold and Yilmaz (2014) to model network connectedness based on volatility spillover from futures prices. Although this approach has been widely used in empirical applications of global stock markets, foreign exchange markets, sovereign bond markets and various assets across countries in a series papers of Diebold and Yilmaz (2012, 2014, 2015a,b) and Demirer et al. (2017), to the best of our knowledge we are among the first research to employ this new approach to study connectedness in commodity futures. In particular, we visualize measures of network connectedness in both static and dynamic settings, and demonstrate volatility-based clustering of commodity futures that match their traditional industrial groupings. Our results show that the energy futures have played a central role with respect to sending volatility to other commodity markets, while the softs and livestocks futures appear to be extremely vulnerable to volatility shocks most of the time.

Our second contribution is a detailed study of commodity connectedness at critical days and periods in the financial markets since 2008. To be specific, we consider two important periods to observe dynamics of volatility shocks and "zoom in" their impacts on markets interactions. The first period comes from the 2007-2009 financial crisis. Our model shows though commodity futures were much closely connected in late 2008, they were lagged in terms of responding to the bankruptcy of Lehman Brothers and stock markets turmoil. Our second period focus on the commodity markets since 2014 when positive supply shocks of crude oil led to a sharp decline in crude oil prices. Surprisingly we find the energy futures have no longer played a central role in the network system and became a net volatility "receivers" one year before price decline. Their impacts on other markets were trivial until crude oil prices became stable in early 2015. The modeling framework we apply seems to be a very promising monitor of markets connectedness in a dynamic context.

The remainder of this paper is organized as follows. Section 3.2 provides a brief outline of VAR model, generalized forecast error variance decomposition and the related network connectedness for range-based volatility. We also discusses the commodities under consideration and how to construct rolling future contracts to meet data requirement for estimating network connectedness in VAR models. Section 3.3 presents our major findings from the full sample static analysis. Section 3.4 shows how to estimate connectedness measures in a dynamic setting and demonstrates interesting results on some critical days/periods of financial markets in the past decade. Section 3.5 concludes and suggests some potential directions for future studies.

3.2 Modeling framework and data

In this section we present the basics of our modeling framework for estimation of connectedness in commodity futures markets. Section 3.2.1 discusses how reduced-form VAR models and generalized variance decomposition can be used to construct static and dynamic connectedness using range-based volatility. Section 3.2.2 introduces the commodity markets in the study and show how we construct the rolling time series to handle various expiry dates of commodity futures.

3.2.1 VAR models for network connectedness of volatility

Following the seminal work of Diebold and Yilmaz (2014, 2015b), we estimate reduced-form VAR approximating models to construct connectedness measures from the H-step ahead generalized forecast error variance decomposition. To be specific, we consider the following N-dimensional VAR(p) model:

$$(3.1)\ \ y_t = \sum_{i=1}^{p} \Phi_i y_{t-i} + \epsilon_t$$

where $\epsilon_t \sim (0, \Omega)$, and Ω is the stationary covariance of ϵ. The related moving average representation is

$$(3.2)\ \ y_t = \sum_{i=0}^{\infty} A_i \epsilon_{t-i}$$

where the $N \times N$ coefficient matrices A_i is supposed to be $A_i = \sum_{j=1}^{p} \Phi_j A_{i-j}$. A_0 is an $N \times N$ identity matrix and A_i=0 for $i < 0$. We transform these moving-average coefficient matrices to obtain variance decompostion such that we can split H-step-ahead forecast error variances of each commodity returns and account for the system shocks in the VAR(p) model.

We do not use the popular variance decomposition such as structural VAR or Cholesky factor since they both requires orthogonalization of VAR shocks, but they assume some other conditions (structural VAR) or depend on the ordering of variables (Cholesky factor). We propose to use the generalized forecast error variance decomposition of Koop, Pesaran, and Potter (1996) and Pesaran and Shin (1998), which is robust to the ordering of variables and does not assume additional conditions, to take correlated shocks into account. In particular, we define variable j's contribution to variable i's H-step-ahead generalized forecast error variance as:

$$(3.3)\ \ \theta^g_{ij}(H) = \frac{\sigma^{-1}_{jj}\sum_{h=0}^{H-1}(e'_i A_h \Omega e_j)^2}{\sum_{h=0}^{H-1}(e'_i A_h \Omega A'_h e_i)}$$

where e_j is the selection vector with j-th element as 1 and 0 elsewhere, and σ_{jj} is the standard deviation of ϵ for variable j*. We proceed to normalize this variance decomposition matrix such that the row sums are one:

$$(3.4)\ \ \dot{\theta}^g_{ij}(H) = \frac{\theta^g_{ij}(H)}{\sum_{j=1}^{N}\theta^g_{ij}(H)}$$

This normalization facilitates interpretation of variance decomposition and provides a directional measure of pairwise connectedness from j to i with predictive horizon H. To simplify notation we can write $C_{i\leftarrow j}(H) = \dot{\theta}^g_{ij}(H)$. It is natural to define net pairwise directional connectedness $C_{ij}(H) = C_{i\leftarrow j}(H) - C_{j\leftarrow i}(H)$. We can also derive aggregate "from" and "to" connectedness measures such that we can investigate the total "influence" an arbitrary variable i exerts or receives in the system of VAR(p) model:

$$(3.5)\ \ C_{i\leftarrow\bullet}(H) = \frac{\sum_{j=1,j\neq i}^{N}\dot{\theta}^g_{ij}(H)}{N}\times 100$$

and

$$(3.6)\ \ C_{\bullet\leftarrow i}(H) = \frac{\sum_{j=1,j\neq i}^{N}\dot{\theta}^g_{ji}(H)}{N}\times 100$$

and we may calculate the net total directional connectedness for variable i simply as (6)-(5):

$$(3.7)\ \ C_i(H) = C_{\bullet\leftarrow i}(H) - C_{i\leftarrow\bullet}(H)$$

*Swanson and Granger (1997) is among the first study that proposes to test structural models of the errors in VAR model. Our approach is built upon the idea of Swanson and Granger (1997) to allow the errors from a VAR model to guide the structural representation of contemporaneous structure.

We can eventually aggregate all "to" or "from" measures and take their means as a system-wide measure of total connectedness:

$$(3.8)\ \ C(H) = \frac{\sum_{i=1}^{N} C_{\bullet \leftarrow i}(H)}{N} = \frac{\sum_{i=1}^{N} C_{i \leftarrow \bullet}(H)}{N}$$

We are especially interested in these measures and will discuss them with more details in next section. Following Diebold and Yilmaz (2015b) we use Garman and Klass (1980)'s approach for intraday range-based volatility that we are going to use in the VAR(p) model:

$$(3.9)\ \ \hat{\sigma}^2 = 0.511(h-l)^2 - 0.019[(c-o)(h+l-2o) - 2(h-o)(l-o)] - 0.383(c-o)^2$$

where h, l, o and c stand for the log of daily high price, low price, opening price and close price respectively. Volatility is always treated as "fear gauge" or sentiment of investors, and we focus on the interdependence of volatility in various commodity futures to explore the transmission mechanism of commodity markets sentiments since it is little studied in the current literature.

3.2.2 Commodity futures data

We consider 20 commodities in the Goldman Sachs Commodity Index (GSCI) that serves as a benchmark for investment in the commodity markets. In particular we have 3 commodities in energy sector (WTI crude oil, Brent crude oil and natural gas), 5 commodities in grains sector (corn, soybean, wheat, soybean oil and rough rice), 6 commodities in softs sector (coffee, cotton, sugar, cocoa, lumber and orange juice), 3 commodities in livestocks sector (feeder cattle, lean hogs and live cattle) and 3 commodities in metals sector (gold, silver and copper).

We obtain commodity futres data from 2nd Jan 1996 to 26th Feb 2016 from Bloomberg terminal. As mentioned in subsection 3.2.1, we need to have daily high price, low price, opening price as well as close price to estimate intraday range-based volatility, which is not provided in the Datastream. Following Christoffersen, Lunde, and Olesen (2014) we construct rolling futures by comparing the trading volumes of the first-nearest to expire contract and the second-nearest to expire contract, and roll to the second contract if its trading volume is greater. We present summary statistics for these volatilities in Table 3.1.

3.3 Static analysis of connectedness

In this section we present empirical results of static estimation of volatility connectedness using the modeling framework in section 3.2.1. We first show in section 3.3.1 the network graph for static measures of connectedness based on all observations in the sample. To investigate interactions between different groups of commodities, section 3.3.2 presents a sector-based connectedness table aggregated and transformed from our static VAR model.

3.3.1 Static connectedness network

We follow Diebold and Yilmaz (2014, 2015b) to estimate a VAR(3) model with predictive horizon $H = 12$ days[†] using all observations in the sample from January 1996 to February 2016. We use network graph to understand the structure of volatility spillover with various nodes and edges to present information in the network. Since there are 380 pairwise directional connections in the model, for simplicity we restrict our graph to the "net" connectedness: any edges between nodes only has one direction, which is estimated by their degree of pairwise connectedness. In our

[†]As indicated by Diebold and Yilmaz (2014) there is no reason why connectedness should be robust to the choice of H, which is solely determined by the purpose of research: when value-at-risk (VaR) is of concern, $H = 10$ is consistent with the requirement of Basel accord. When the model is used for portfolio optimization then H should correspond to the specific rebalancing period. In this study, robustness check shows that varying orders and predictive horizon in the VAR models only has little impacts on our results.

Table 3.1: Summary statistics of futures volatility

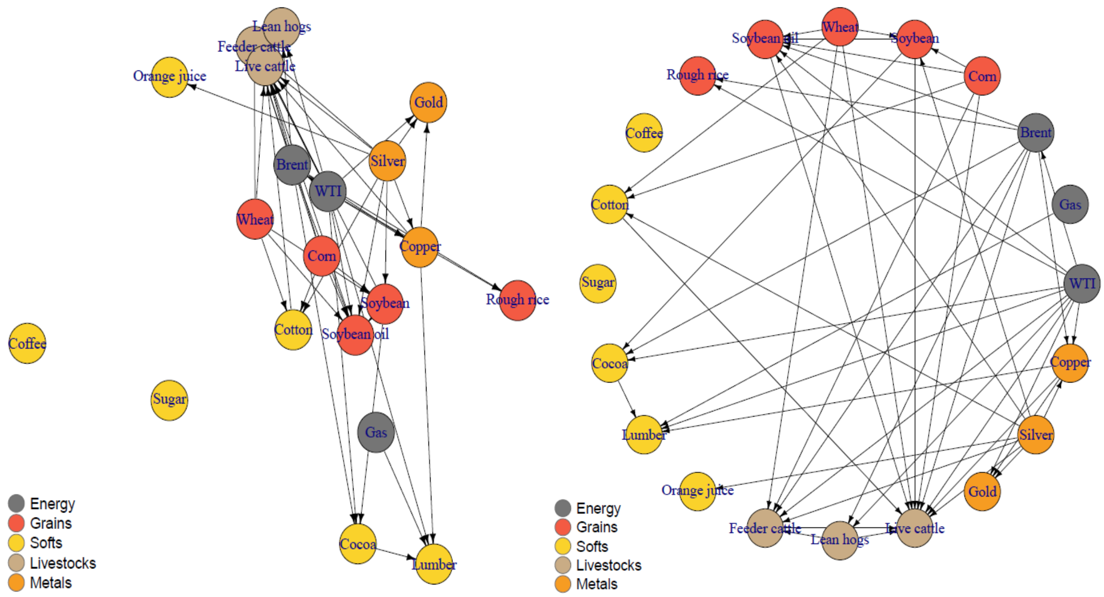
Commodity	Mean	Minimum	Maximum	Std. Dev.	Skewness	Kurtosis
Energy						
WTI	0.0195	0.0014	0.1240	0.0101	2.1666	11.1976
Gas	0.0274	0.0023	0.2282	0.0139	2.4299	18.8256
Brent	0.0175	0.0024	0.1126	0.0095	2.1862	11.5824
Group average	0.0215	0.0039	0.0930	0.0086	1.6316	8.0046
Grains						
Corn	0.0138	0.0011	0.0753	0.0074	1.7491	8.1731
Soybean	0.0126	0.0027	0.0713	0.0065	2.0921	10.3764
Wheat	0.0166	0.0045	0.1582	0.0080	2.7990	28.2995
Soybean oil	0.0128	0.0020	0.0685	0.0062	2.0651	10.8481
Rough rice	0.0121	0.0003	0.0773	0.0070	1.8637	10.2342
Group average	0.0136	0.0047	0.0494	0.0051	1.7490	8.0772
Softs						
Coffee	0.0183	0.0013	0.0836	0.0087	1.6970	7.6958
Cotton	0.0131	0.0007	0.0703	0.0072	1.7646	8.4189
Sugar	0.0163	0.0011	0.0898	0.0086	1.6376	8.2045
Cocoa	0.0143	0.0025	0.0885	0.0074	1.9649	10.9656
Lumber	0.0141	0.0008	0.0619	0.0064	1.3860	6.7022
Orange juice	0.0156	0.0008	0.1045	0.0098	2.0245	9.9737
Group average	0.0153	0.0065	0.0414	0.0040	1.1162	5.3984
Livestock						
Feeder cattle	0.0072	0.0006	0.0327	0.0037	1.6726	7.7129
Lean hogs	0.0121	0.0003	0.0915	0.0058	2.1796	15.6886
Live cattle	0.0075	0.0009	0.0303	0.0034	1.4772	6.5267
Group average	0.0090	0.0026	0.0417	0.0032	1.6322	8.8330
Metals						
Gold	0.0095	0.0009	0.0893	0.0062	3.0343	22.3146
Silver	0.0168	0.0019	0.1743	0.0105	3.1318	24.3074
Copper	0.0151	0.0013	0.0938	0.0088	2.5938	13.9563
Group average	0.0138	0.0038	0.1068	0.0072	2.8996	18.9009

Notes: Range-based volatility of 20 commodity futures from Jan 2, 1996 to Feb 26, 2016. The continuous futures series are constructed by rolling to the second-nearest to expire contract when its trading volume is greater than the trading volume of the first-nearest to expire contract.

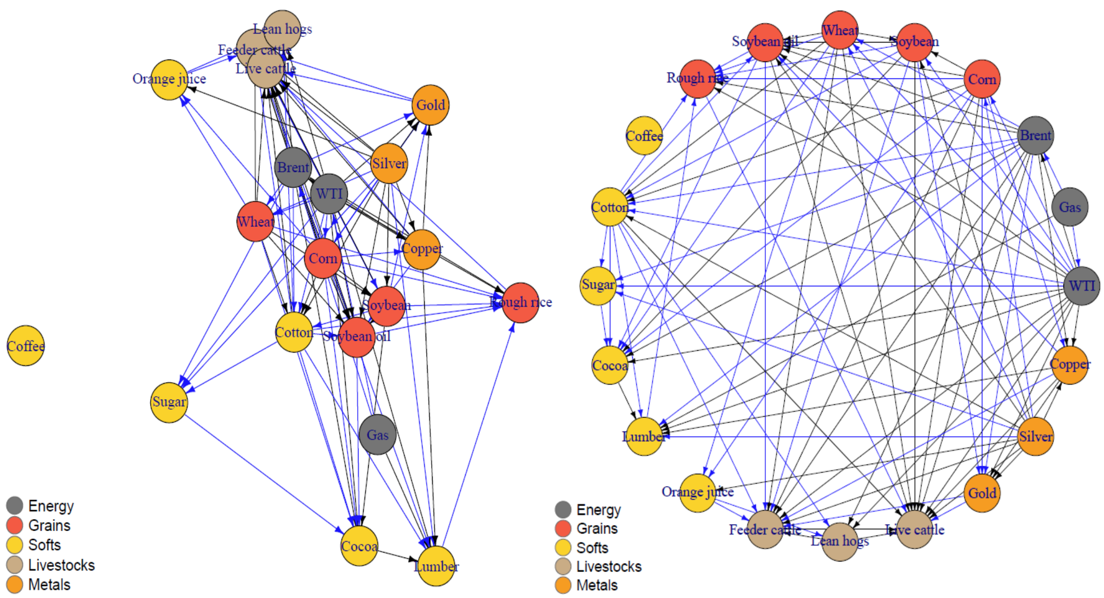
network graph, node color indicates group membership of commodity; edge color indicates percentile of connectedness strengths; edge arrow indicates pairwise direction of connectedness. Node location is determined by the ForceAtlas2 algorithm of Jacomy et al. (2014), which searches for a steady state when the repelling and attracting forces are equal. To be specific, in our text this algorithm assumes nodes repel each other while the links attract and connect the nodes according to the average net pairwise directional connectedness. As a result, distance between nodes decreases as their degree of connectedness increases.

Figure 3.1 shows the network graph of static volatility connectedness in the commodity markets. Nodes in different colors represents industrial grouping of commodities. Figure 3.1 (a) and (c) demonstrate network connectedness where the node location is determined by the ForceAtlas2 algorithm. To better capture “senders” and “receivers” of volatility Figure 3.1 (b) and (d) contain the same network information but the nodes are placed in a circle. Figure 3.1 reveals evidence of commodities clustering but the strength of clustering varies in different commodity groups. In Figure 3.1 we do not report all 190 pairwise directional edges in the network. To emphasize on the most important connections among commodities we instead only show the 25^{th} and 50^{th} percentiles of pairwise connectedness measures in black and blue edges respectively.

In Figure 3.1 (a) and (b), the respective locations of commodities in grains, livestocks and metals sectors show clear evidence of sector clustering. In energy sector, there is high pairwise connectedness between Brent and WTI, but natural gas seems quite separated from the oil commodities. In softs sector, commodities like orange juice, coffee and sugar are located far away from other members in the same sector. Cocoa and lumber, however, are close to each other, but they are located futher away from the bulk of other commodities. Our graph plot reveals that cotton



(a) Upper 25% net spillover (b) Upper 25% net spillover in circle layout



(c) Upper 50% net spillover (d) Upper 50% net spillover in circle layout

Notes: we show the strongest 50% connections among the pairs of 20 commodities. Nodes in grey, red, gold, tan and orange represents commodity groups of energy, grains, softs, livestocks and metals. Black and blue edges correspond to the 25^{th} and 50^{th} percentiles of all net pairwise directional connections in the static analysis.

Figure 3.1: Net pairwise volatility spillover in the full sample from Jan 1996 to Feb 2016

and orange juice, though both belong to the softs sector, are respectively next to the grains sector and livestocks sector, suggesting they are more connected with commodities in these sectors. The node locations of Brent and WTI relative to the bulk of other commodities imply they play a central role in volatility spillover. This is not surprising since the prices of most non-energy commodities heavily depend on oil price through transpotation. Therefore non-energy markets are more vulnerable to the prices shocks and volatility from oil markets.

From Figure 3.1 (b) and (d) we can easily find some obvious large "senders" and "receivers" of volatility in commodity markets, which are in the center of Figure 3.1 (a) and (c) as the ForceAtlas2 algorithm are symmetric for "to" and "from" measures when it determines the location of nodes. First of all, we find that Brent and WTI are major senders of volatility in the network and have a wide range of impacts across and within sectors. Brent and WTI have spillover effects on similar commodities, partly due to the fact that their prices are driven by the same macroeconomic and industry-wide shocks. Secondly, we identify feeder cattle, live cattle, soybean oil, cocoa and lumber, all from agricultural markets, as large receivers of volatility in the network. They share one important characteristic: the volatilities they get come from commodities in all sectors. This finding makes sense since agricultural products have their own production circles and may not be as sensitive to the macroeconomic shocks as energy markets. Thus they appear to be lagged in the volatility network and led by oil futures.

3.3.2 Sector-based connectedness table

Although Figure 3.1 is useful for revealing important pairwise directional connectedness in the network, it is not very informative to show spillover effects for all pairwise relations. Another concern is though we observe evidence of sector clustering, we still don't figure out the interactions of volatility spillover among different sectors. Since there are 20 commodities in our model, it is not practical to present all of them with a 20×20 connectedness table. Therefore we cluster all commodity futures based on the traditional industrial groupings and aggregate connectedness measures accordingly. We eventually arrive at a 5×5 connectedness table below:

Table 3.2: Full-sample connectedness table

To\From	Energy	Grains	Softs	Livstocks	Metals	**Sum of From**
Energy	30.11	0.53	0.23	0.49	1.37	32.74
Grains	1.05	16.47	1.13	0.46	2.12	21.23
Softs	0.66	1.3	14.69	0.24	0.91	17.79
Livestocks	2.31	1.16	0.45	26.90	1.30	32.11
Metals	1.90	1.89	0.67	0.51	26.43	31.40
Sum of To	36.04	21.33	17.17	28.59	32.13	**34.44**
Net	3.30	0.11	-0.62	-3.52	0.73	

Note: The sample is from Jan 2, 1996 to Feb 26, 2016, and the predictive horizon is 12 days. The ij-th entry of this 5×5 matrix represents pairwise directional connectedness from j to i. The rightmost column is the sum of volatility spillover received by different sector. The bottom "Sum of To" row is the sum of volatility spillover from any sectors to the others. The bottommost "Net" row is simply the difference between "Sum of To" and "Sum of From". The intersection of "Sum of From" and "Sum of To" is the total connectedness measure based on individual "Sum of From" in the VAR model system.

From Table 3.2 we can find some notable features of connectedness in commodity markets. For example, the diagonal element is always much higher than the other elements in the same row, suggesting most of the volatility shocks come from the commodities in the same sector. It seems that commodity markets are quite segmented since at least 80% of the sum sends or receives by any sector is actually from the same sector. Energy sector has the highest "Net", implying that it has the largest influence on the other four sectors, while livestocks sector has the lowest "Net", indicating that its influence on the other commodities is very trivial and is basically a large recipient of volatility in the futures markets. The total connectedness of the system is only 34.44%, which is relatively small compared to the same measure (78.3%) in Diebold and Yilmaz (2014), who study volatility spillover of 13 major U.S. financial firms from May 1999 through April 2010. However, since Table 3.2 is based on the full sample, our static VAR model could not capture significant variations in the volatility of commodity markets and potential structural changes in the VAR estimates. As a next step we attempt to reexamine these connectedness measures in a dynamic framework, such that we can observe the interactions of volatility spillover among futures markets during the period of financial distress or other significant moments in financial markets.

3.4 Dynamic analysis of connectedness

In this section we discuss how to estimate dynamic VAR models, derive time-varying spillover and use them to monitor and describe evolution of volatility connectedness in commodity markets during the 2007-2009 financial distress and the recent downward movement of commodity prices since mid-2014. Section 3.4.1 shows it is important to consider dynamics in the total connectedness measures, which are extremely sensitive to market sentiment and should be examined carefully across financial cycles. Section 3.4.2 presents time-varying volatility spillover across sectors. Section 3.4.3 describes volatility co-movements in futures markets during the financial crisis in 2008. Section 3.4.4 analyzes the impacts of supply shocks from energy futures on other commodities.

3.4.1 Time-varying total connectedness

We still employ the VAR(3) model and 12-day-ahead forecast error for generalized decomposition in the dynamic analysis, but now estimate with a 252-day[‡] (all trading days in a year) rolling-window to model dynamics in the volatility connectedness, as suggested by Diebold and Yilmaz (2014, 2015a,b) and Demirer et al. (2017). The total connectedness plot is showed in Figure 3.2 below. Notice that on Sep 28 1999 there was an extraordinary spike in volume, volatility and price occurred in the gold options market at the New York Mercantile Exchange (NYMEX), which led to an ad hoc increase in the total connectedness index. This is due to the unexpected announcement of 15 European central banks that claimed a surprise five-year moratorium on all new sales of gold held in official reserves. The connectedness index is quite low and stays around 45% after this event till 2007. We can see dramatic increase in total connectedness during the 2007-2009 financial crisis but it has decreased since late 2009. The peak of total connectedness is in the week of Oct 26th, 2008 which is one month later than the Lehman Brothers' bankruptcy on Sep 15th, 2008. It appears that during financial crisis stock markets have led commodity markets since Diebold and Yilmaz (2014, 2015b) who study the stock returns of financial institutions in the U.S. and Europe find this dynamic connectedness has reached its peak right after the bankruptcy of Lehman Brothers. The most influential event in the commodity markets since 2014 is perhaps the sharp decline in crude oil prices, but Figure 3.2 shows that this decline does not imply increasing connectedness in all 20 commodity markets.

Recall that the static measure of connectedness in our full-sample unconditional analysis is only 34.44%, which is apparently much smaller than the ranges of dynamic

[‡]We have used 150 days, 200 days and 300 days as alternative rolling-window and the results are still similar.

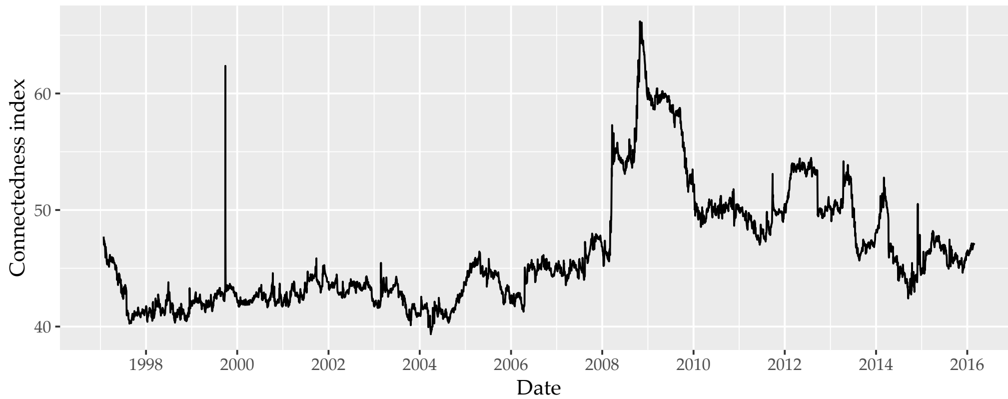


Figure 3.2: Dynamic total connectedness with a 252-day rolling window and 12-day-ahead predictive horizon for variance decomposition

measures most of the time. This suggests that static analysis may underestimate connectedness and it is critical to take recent dynamics into consideration. Notice that this remarkable difference between dynamic and static measure of connectedness, however, is not found in existing literature which focuses on global bank network connectedness (Diebold and Yilmaz 2014, 2015b; Demirer et al. 2017). We suspect the dynamic measures of connectedness for commodity markets are more sensitive to the length of rolling estimation window.

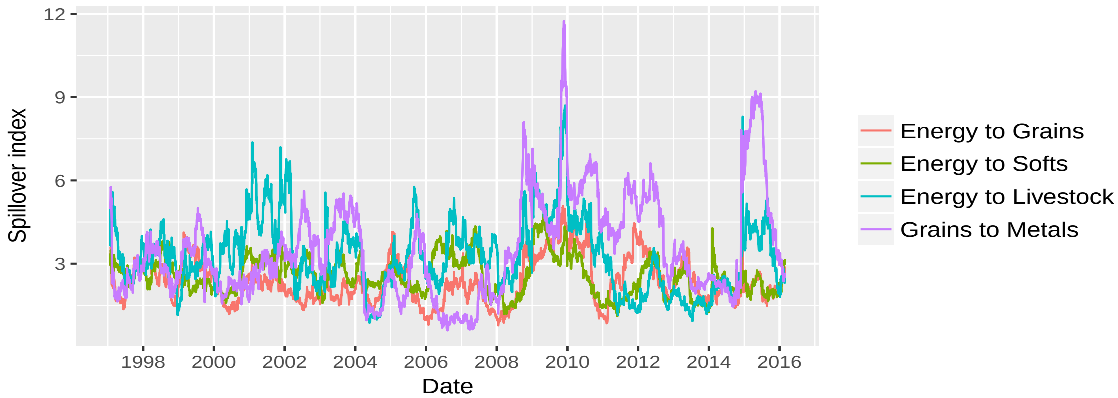
3.4.2 Time-varying connectedness across sectors

In this subsection we explore the dynamic measures of volatility connectedness at sector level. Again due to the fact that we have 20 commodities and cannot present all 380 pairwise directional connectedness in a very concise way, we omit the diagonal-elements in the dynamic connectedness table to focus on cross-sector connectedness. We can see from Figures 3.3 and 3.4 that cross-sector volatility spillover have increased dramatically during the 2007−2009 financial crisis, but they soon returned to the pre-crisis level after distress period. Surprisingly, recent downward price movements of energy markets do not have obvious spillover effects on livestocks, softs and metals markets as usual until crude oil prices became stable in early 2015. These figures show that directional pairwise volatility spillovers are quite volatile even when the economy is expanding, posing challenge on some recent research that assert commodity markets are only highly integrated during financial distress. Figure 3.4 (b) implies the metals futures are another major senders of volatility, but their spillovers effects are much less volatile than their energy counterparts.

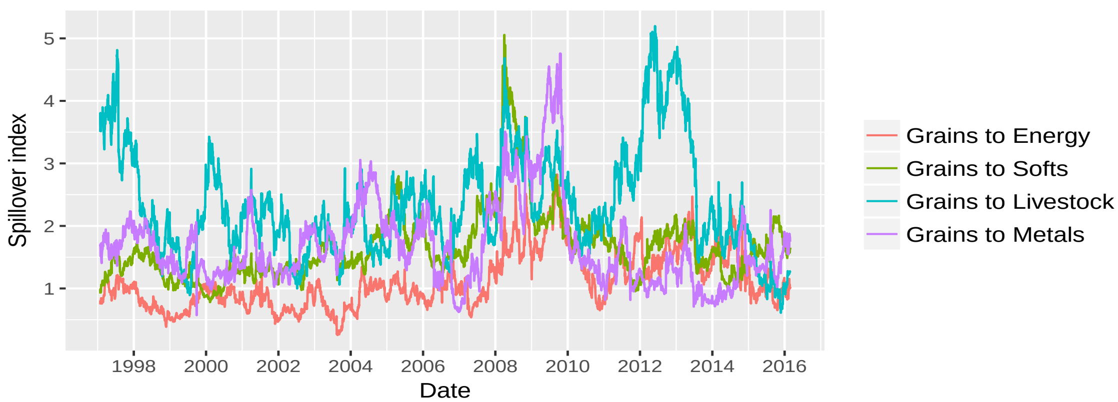
3.4.3 Market connectedness during financial crisis

As we have mentioned in Section 3.4.1, the dynamic total connectedness has increased to the highest level on Octorber 29, 2008 in our data. We now discuss the details of connectedness measures and network graph for this particular date. As we will see below, there are striking differences in connectedenss measures between the static model and the dynamic model during financial stress. First of all, the dyanmic total connectedness reached the peak at 66.19%. Recall that in our static model this measure is only 34.44%, implying the total connectedness among commodity futures has almost doubled when the financial system was going down. This is not uncommon since most of the literature on risk management has documented financial markets tend to crash together, not to boom together. As previous literature has largely focused on prices and returns, our finding provides another new but consistent perspective of market connectedness through volatility co-movements.

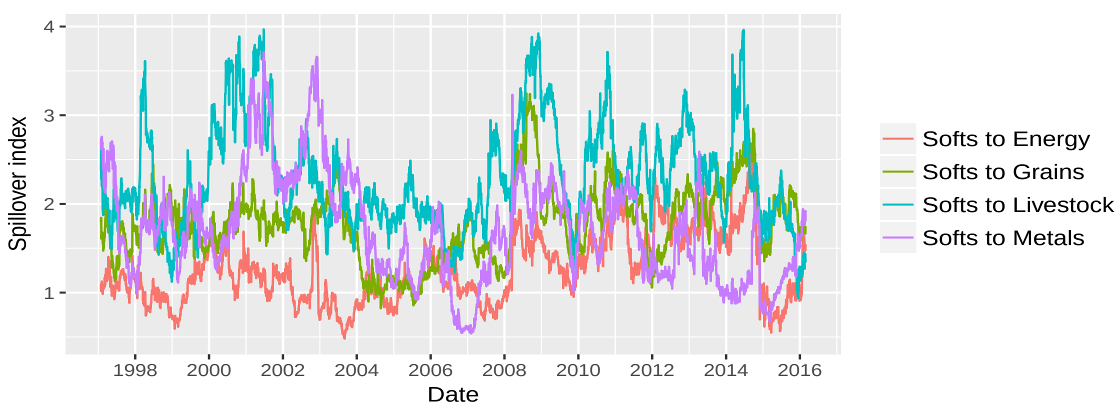
Secondly, the network graph in Figure 3.5 shows the commodity markets have been more connected with each other than in the static one in Figure 3.1. For example, all markets except lean hogs have clustered and there are edges to connect every node in the network. Sector clustering, however, is not as obvious as we have seen in the static network. In Figure 3.1 (b) and (d) it is very clear that energy and metals commodities are the major senders of volatility, but the node locations



(a) Spillover from energy to other commodities

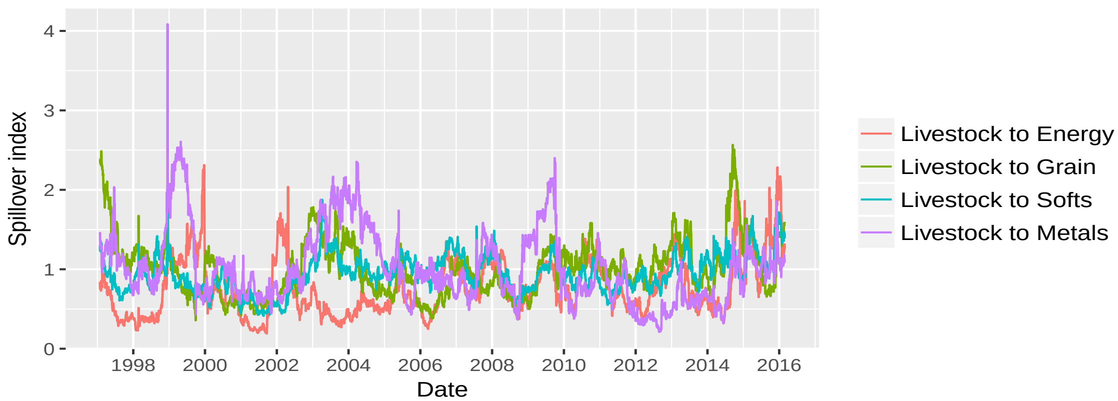


(b) Spillover from grains to other commodities

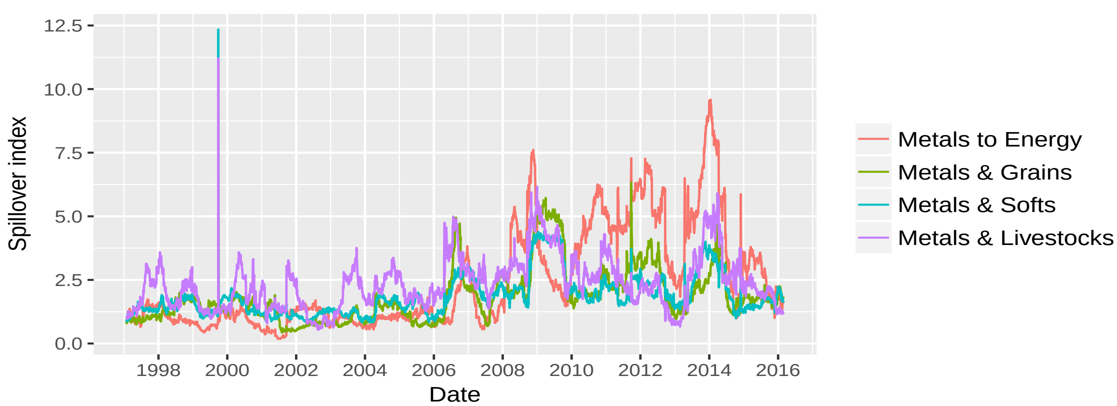


(c) Spillover from softs to other commodities

Figure 3.3: Volatility spillover from energy, grains and softs to other commodities

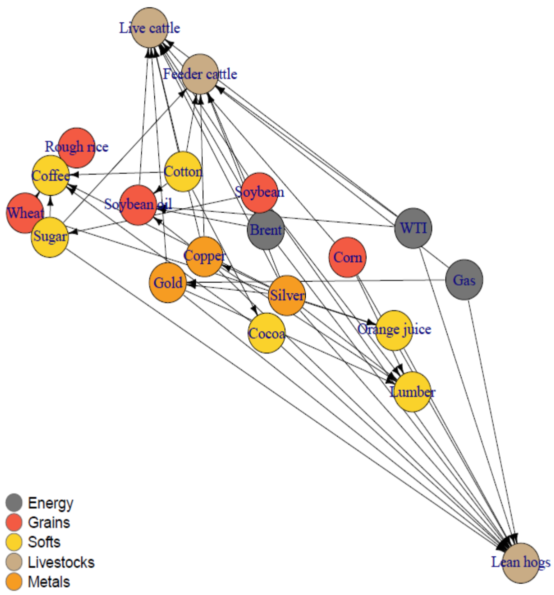


(a) Spillover from livestock to other commodities

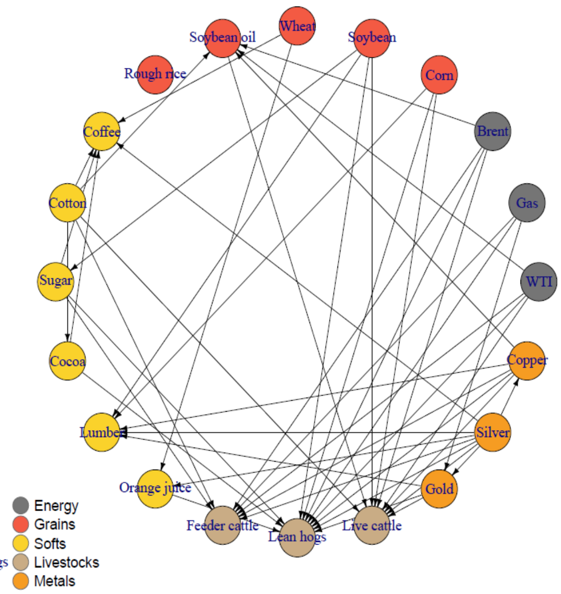


(b) Spillover from metals to other commodities

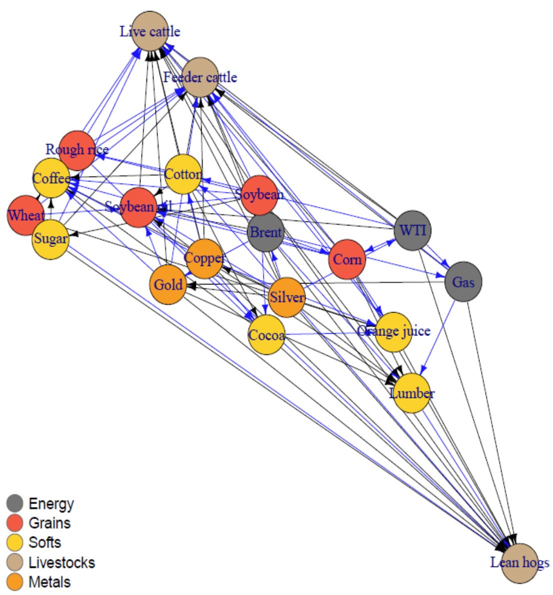
Figure 3.4: Volatility spillover from livestock and metals to other commodities



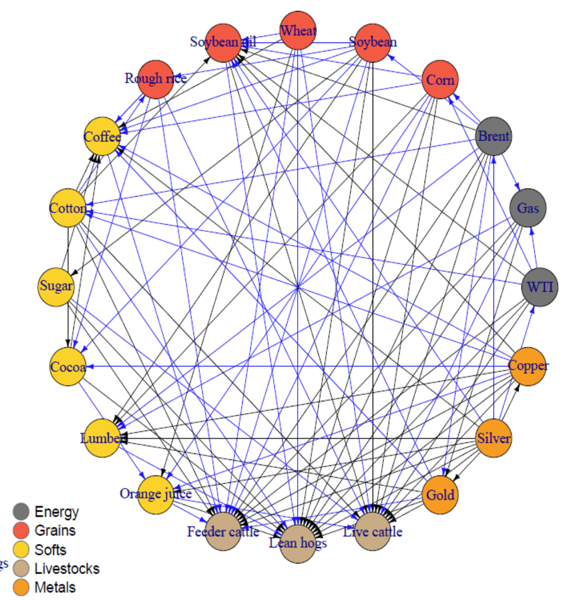
(a) Upper 25% net spillover



(b) Upper 25% net spillover in circle layout



(c) Upper 50% net spillover



(d) Upper 50% net spillover in circle layout

Notes: we show the strongest 50% connections among the pairs of 20 commodities. Nodes in grey, red, gold, tan and orange represents commodity groups of energy, grains, softs, livestocks and metals. Black and blue edges correspond to the 25^{th} and 50^{th} percentiles of all net pairwise directional connections in the dynamic analysis.

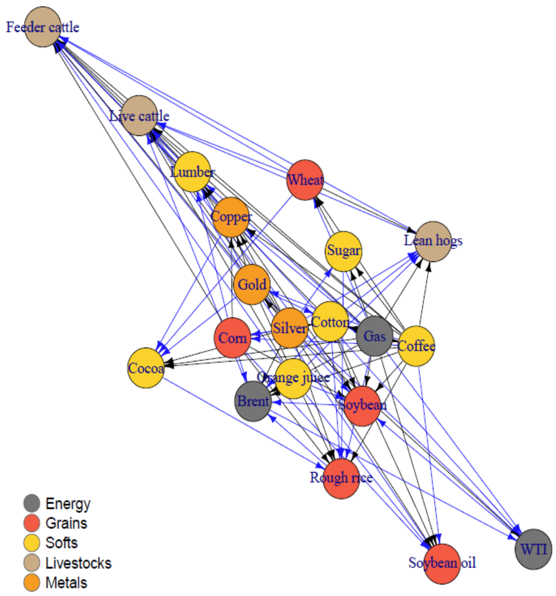
Figure 3.5: Net pairwise volatility spillover during financial crisis on Oct 29, 2008

of energy futures are no longer in the center of all markets. Livestocks and most of the futures in grains and softs sectors seem to be much more vulnerable to the volatility spillovers during the financial crisis. To summarize, Figure 3.5 presents a clear description of markets interactions when demand shocks from macroeconomy have influenced all markets at the same time.

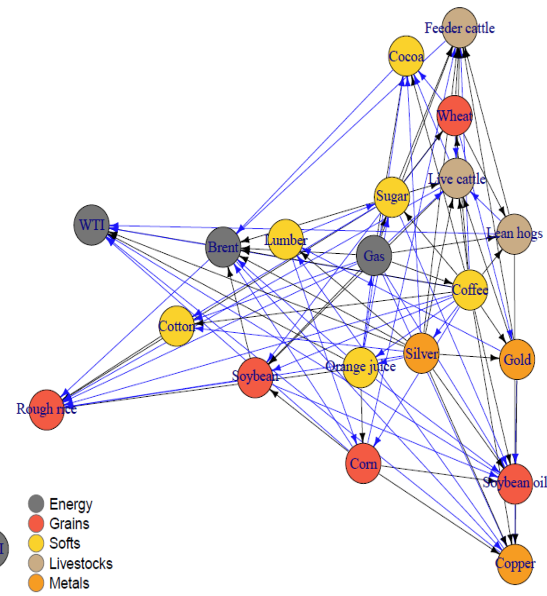
3.4.4 Market connectedness with crude oil supply shocks

The sharp decline in crude oil prices since mid 2014 is probably the most important event in the commodity markets in recent years. Due to the unexpected expanding production of crude oil in the U.S., WTI price has dropped from $107.26 per barrel on June 20, 2014 to $45.59 per barrel on January 23, 2015. We note that the bottom of WTI price is only around $50 per barrel during the financial crisis. Since crude oil has the largest share of trading volume and great impact on other futures, it would be interesting to know the directions and magnitude of volatility spillovers when this supply shocks hit the markets. We select four dates, June 20 2014, September 5 2014, November 28 2014 and January 23 2015, which are almost evenly distributed along the crude oil price decline process, to monitor volatility co-movements among commodity futures and present the results in Figure 3.6 .

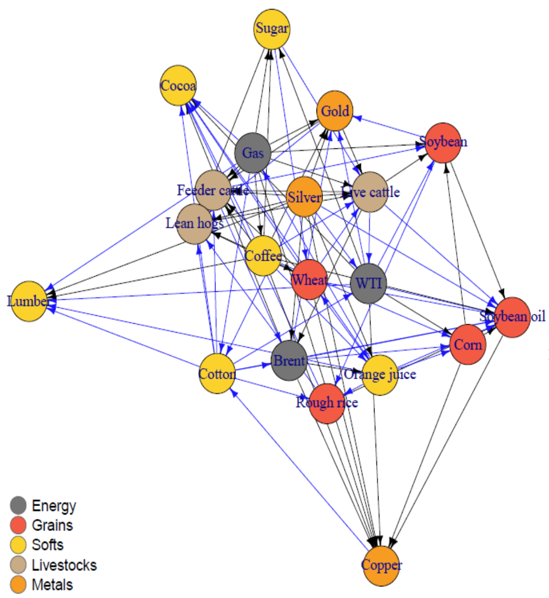
We document two interesting findings in Figure 3.6. We first focus on the directions of net volatility spillovers "from" and "to" crude oil futures. Figure 3.6 (a) shows that at the very beginning of price decline process, commodity markets appear to be "abnormal" since Brent and WTI, which are large senders of volatility in the static model and dynamic model during financial crisis, have already become large receivers of volatility on June 20, 2014, when crude oil price was still high. These unusual movements of volatility suggest other commodities may have already received supply shocks of crude oil due to price discoveries in the futures markets.



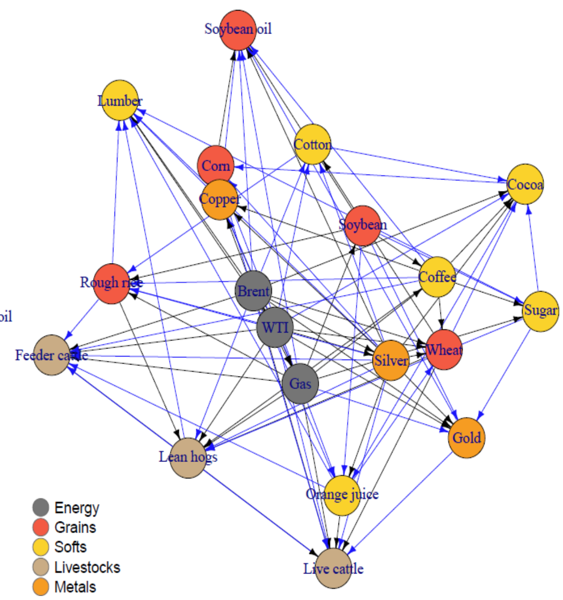
(a) Network on June 20, 2014.



(b) Network on Sep 5, 2014.



(c) Network on Nov 28, 2014.



(d) Network on Jan 23, 2015.

Notes: we show the strongest 50% connections among the pairs of 20 commodities. Nodes in grey, red, gold, tan and orange represents commodity groups of energy, grains, softs, livestocks and metals. Black and blue edges correspond to the 25^{th} and 50^{th} percentiles of all net pairwise directional connections in the dynamic analysis.

Figure 3.6: Net pairwise volatility spillover during energy supply shock from mid-2014 through early 2015

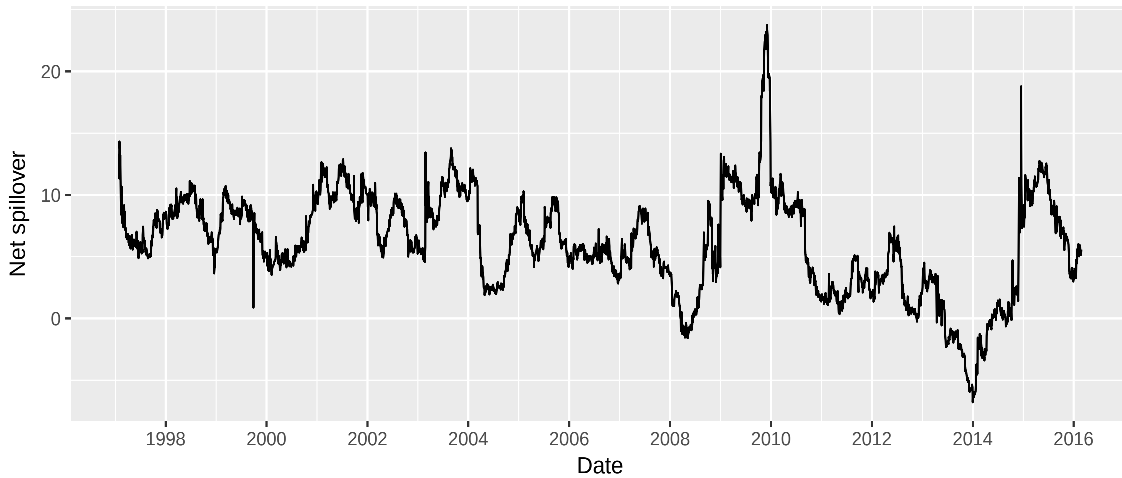


Figure 3.7: Net volatility spillover of energy markets

When the supply shock unfolded on September 5, 2014, Brent and WTI received much more volatilities and barely sent any to other markets, as we can see in Figure 3.6 (b). At the final stage of price decline on January 23, 2015, the network returned to "normal" status as Brent and WTI again became large senders in the markets.

Second, the respective locations of Brent and WTI also show their impacts on other markets have already decreased when the supply shock has not yet arrived. For example, WTI is far away from the cluster of nodes in Figure 3.6 (a) and (b), and the energy markets have not returned to the center of commodity markets until the end of price decline in January 2015. We think the modeling framework is informative in the discovery process of volatility spillovers and may serve as a very useful monitor of commodity futures markets in the real time.

Figure 3.7 presents the net volatility spillover of energy markets, which is simply the difference between "To" and "From" measures of all three futures in the energy sector. We can see the net spillover values are always above zero before early 2013, implying that energy futures are large volatility senders in the commodity markets. It is interesting to observe that energy sector became a net receivers since early 2013, which is almost one year before sharp decline in energy prices. The net spillover reached bottom at the beginning of 2014 and gradually went up since then and finally return to a net sender before 2015. It seems that during its price decline period energy sector did not contribute much spillovers to the other commodity markets, and once the crude oil price became stable energy markets resumed its traditional role in the commodity markets.

3.5 Concluding remarks

In this paper, we characterize static and dynamic volatility spillovers in commodity futures markets from 1996 to 2016. In our static analysis we show that commodity futures are much more connected with those in the same industrial group, and energy commodities play a central role and have greater influence on the other futures. In our dynamic analysis we find the total connectedness measure of futures markets has dramatically increased during the financial distress in 2007-2008. This measure of market connectedness based on volatility spillover, however, has decreased gradually after the crisis, despite the fact that there is a sharp decline in crude oil prices since 2014. We demonstrate energy markets have much higher degree of spillovers to other markets in the dynamic setting, which is consistent with the results from static model. Network graphs and dynamic measures of connectedness on some critical dates exhibit: 1) all markets were closely linked to each other during financial crisis; 2) price shocks from crude oil futures weakended their spillovers effects on other commodities until the shocks have disappeared in early 2015. Our empirical results suggest the modeling approach for volatility spillovers of Diebold and Yilmaz (2014) could be used as an informative tool to provide co-movements signal and monitor markets connectedness in the real time, which has vast applications in risk management and portfolio construction.

We discuss several interesting directions to improve and extend the current method for future studies. First, Kilian (2009) shows that depending on the driven demand or supply shocks oil price may have different impacts on macroeconomy and commodity markets using a structural VAR model. To deepen our understanding in volatility spillovers driven by different mechanisms across financial and macroeconomic cycles, we wonder if we may in the spirit of Kilian (2009) estimate a macro-finance regression model that links macroeconomy to the dynamics of volatility spillovers and market connectedness. Mixed-data sampling regression models (MIDAS) of Ghysels, Santa-Clara, and Valkanov (2004) seems to be a very promising approach for tackling this challenge. Second, it would be great to develop a framework for guiding financial market participants in hedging and risk managment based on various connectedness measures in the model. Last but not least, the practical value of dynamic analysis from the current modeling framework heavily depends on rolling window estimation of VAR model, which is not necessarily robust to the choice of window length. A flexible time-varying approach that allows for structural breaks in VAR model is the key to precise measures of markets connectedness across financial/macroeconomic cycles. Although we have verified robustness of dynamic VAR model results using various rolling-window in our analysis, it is appealing to use a data-driven approach to select an optimal rolling window. A recent paper by Inoue, Jin, and Rossi (2017) proposes a promising idea to determine the optimal length of rolling window, but it has not been extended to accommodate latent variables (e.g., volatility, connectedness measures) which are of great interest in the current study. We conclude by raising these concerns and hope they can be extended to a wider range of future studies.

4. SPOUSAL DEPENDENCE AND INTERGENERATIONAL TRANSMISSION OF BODY MASS INDEX: EVIDENCE FROM CHINA

4.1 Introduction

Obesity is a common risk factor associated with many chronic diseases, and its increasing prevalence has imposed tremendous financial burden on countries undergoing rapid economic development. Although there are numerous studies on the undesirable economic and health consequences of obesity in developed countries, academic research is relatively silent on its impacts on developing countries. But this does not imply it is not an important topic in these countries. According to a recent survey (Ng et al. 2014) conducted by the Institute for Health Metrics and Evaluation at the University of Washington, China has the second largest obese population (more than 62 million) in the world. According to the Report on the Nutrition and Chronic Disease Status of Chinese Residents by the National Health and Family Planning Commission of the Peoples Republic of China (2014), the overweight and obesity rates of Chinese adults are 30.15% and 11.9% respectively in 2014, significantly increased from 22.85% and 7.1% in 2002. Body mass index (BMI), defined as height/weight^2 (meters/kg^2), is a common measure to classify overweight and obesity. Using nationally representative household surveys, we depict an increasing trend of obesity and overweight percentages of Chinese adults from 1991 to 2011 in Figure 4.1 below.

A large number of studies show that overweight and obesity are associated with family environment and genetic traits (Vogler et al. 1995; Philipson and Posner 1999; Jeffery and Rick 2002; Sacerdote 2004; Lakdawalla and Philipson 2009; Gao, Zhang, and Wu 2015; Gao and Shen 2016). For example, Jeffery and Rick (2002) report

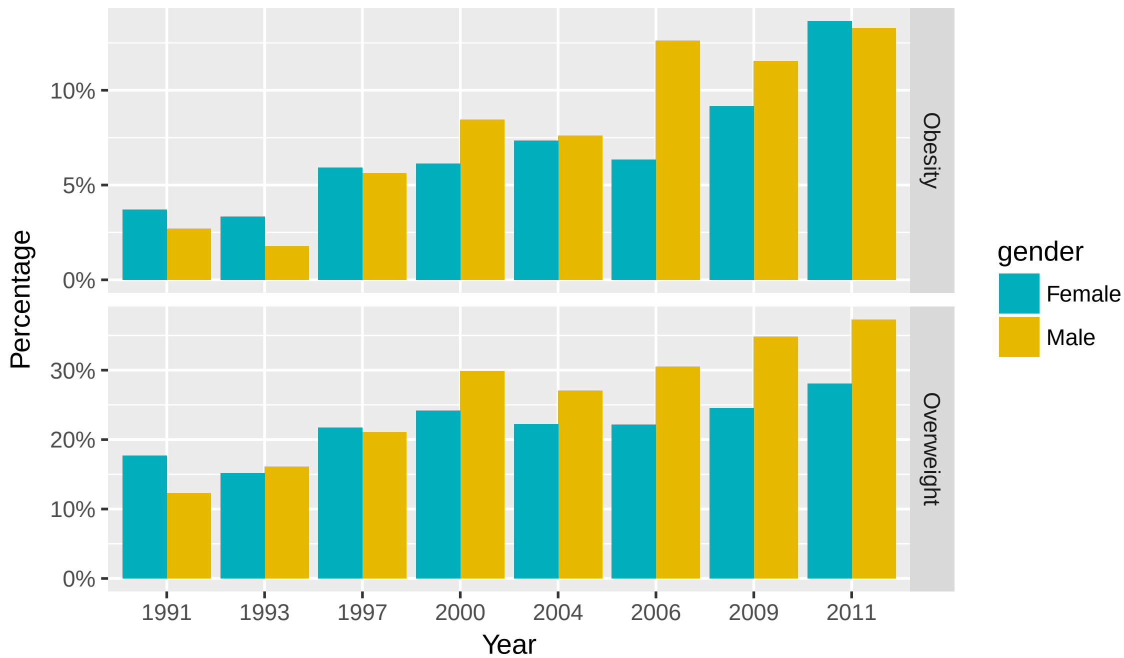


Figure 4.1: Percentages of overweight and obese Chinese adults over time

that spousal correlation in BMI ranges from 0.1 to 0.2. Positive spousal correlation in BMI can be attributed to: (1) assortative matching in selection of spouse, and (2) family environment, including exercise frequency, dietary habits and household income, which are usually common and shared by couples. Using a novel copula model Gao, Zhang, and Wu (2015) find that the intergenerational BMI dependence is generally asymmetric and stronger for females. Intergenerational transmission of BMI arises for two reasons: (1) genetic traits, which are shared by biological parents and children; (2) family environment (where and how parents raise their children). Existing literature indicates that family environment plays an important role in both spousal and intergenerational correlation in BMI. For instance, Anderson, Butcher, and Levine (2003) suggest that if mothers work for long hours their children are more likely to be obese or overweight. Taveras et al. (2005) find that overweight and obesity are negatively related to the frequency of having family dinners.

Some recent studies consider weight/height data augmented with socioeconomic variables (employment status, income, insurance, etc.), demographic information (age, gender, education, residents' area, etc.) and behavioral variables (smoking, exercise frequency, etc.) to help explain observed variations in obesity and overweight (Price, Reed, and Guido 2000; Wilson 2002; Chou, Grossman, and Saffer 2004; Classen and Hokayem 2005; Mamun et al. 2005; Abrevaya and Tang 2011; Cohen et al. 2013; Chen, Liu, and Wang 2014; Gao and Shen 2016). For example, Abrevaya and Tang (2011) use a large micro dataset in the United States with information on husbands, wives and grown children to explore familial BMI relationship and determination of weight status. They find that household income affects husband's and wife's BMI differently; parental BMI and smoking behavior serve as significant predictors for grown children's BMI. Gao and Shen (2016) explore a Chinese data and find different determinants of BMI for urban and rural residents. In particular, they suggest that BMI is correlated with gender, age, labor intensity, drinking and eating habits among urban residents, and with income, number of siblings and eating habits among rural residents.

This paper attempts to analyze the increasing prevalence of obesity and overweight in China over a span of two decades. We make two primary contributions to the current literature. First, we explicitly model familial relationship of BMI in different areas of China. Although Gao and Shen (2016) also consider separate models for Chinese urban and rural residents' BMI, they do not take into account spousal BMI and parental BMI in their analysis despite the abundant evidence on BMI transmission within family. Secondly, our analysis explores the dynamics of obesity and overweight determinants for Chinese in different periods. The importance of dynamics of obesity determinants is demonstrated by Philipson and Posner (1999) and Lakdawalla and Philipson (2009), who argue that agricultural and technological innovations contribute to the increase in overweight and obesity. As China has experienced a rapid economic growth and numerous agricultural innovations since early 1990s, one may naturally conjecture that the determinants of obesity might have evolved during this period. Unlike studies that rely on cross-sectional data (e.g., Abrevaya and Tang (2011) and Gao and Shen (2016)), we utilize nationally representative data that span two decades since 1991. This longitudinal sample allows us to model time varying impacts of various contributing factors on BMI during the sample period.

Our results provide strong evidence on BMI transmission within family. For example, an individual's BMI is found to have a significant and positive impact on the BMI of his/her spouse, though this impact has decreased in the recent decade. Intergenerational transmission of BMI is evident: parental BMI is the most important predictor for children's BMI. We also find that individual BMI depends on socioeconomic and demographic characteristics. In particular, income effect is positive for men's BMI while employment status has a negative effect on women's BMI. However, these characteristics are not as informative for children's BMI, for whom only education attainment is found to be a negative predictor for younger women. Lastly, we identify significant education impact on couples' BMI using structural regressions with a common couple effect, suggesting that some common factors influencing both husband and wife are probably omitted.

The remainder of this paper is structured as follows. Section 4.2 briefly introduces some background information and provides summary statistics of the household survey data used in our analysis. Section 4.3 presents the models and estimation results. Specifically, section 4.3.1 and section 4.3.3 consider proxy-variable regressions for spousal BMI and children BMI, while section 4.3.2 investigates couple effects using correlated random-effects regressions for spousal BMI. The last section concludes.

4.2 Data

This study uses data from the China Health and Nutrition Survey (CHNS) that is an ongoing international collaborative project between Carolina Population Center at the University of North Carolina at Chapel Hill and the National Institute for Nutrition and Health (NINH) at the Chinese Center for Disease Control and Prevention (CCDC). This project is designed to measure the impacts of health, nutrition and family planning policies implemented by national and local governments, and examine how the economic transformation of China affects the health and nutritional condition of the Chinese population. The survey was conducted in 12 provinces (Guangxi, Guizhou, Heilongjiang, Henan, Hubei, Hunan, Jiangsu, Liaoning, Shaanxi, Shandong, Yunnan and Zhejiang) and 3 national central cities (Beijing, Chongqing and Shanghai). The initial round of survey began in 1989; detailed information pertinent to this study was collected in eight subsequent rounds in 1991, 1993, 1997, 2000, 2004, 2006, 2009 and 2011. We therefore focus our analysis on these eight rounds. To capture dynamics in BMI determinants during the sample period and at the same time to avoid yearly sampling variations, we cluster our samples into the 1990s group (1991, 1993, 1997, 2000) and the 2000s group (2004, 2006, 2009, 2011).

CHNS conducts surveys at the household level, collecting information regarding individual household members such as age, education level, height, weight, employment status, smoking behavior and health insurance coverage. Also available are some household common characteristics, such as total annual income (in RMB), child status (has child or not) and region (urban or rural). The key variable in our study, BMI, is measured as weight/height^2 (kg/m^2) in the typical way. Notice that the recommended classification of obesity and overweight is different for the Chinese population because asians tend to have higher body fat than whites of same age and BMI (Potts 2003). In particular BMI values between 24 to 28 are classified as overweight and those greater than 28 as obese*.

We present summary statistics of our sample in Table 4.1. Sample averages and standard deviation of non-indicator variables, as well as percentages of indicator variables are reported separately in the upper and lower panel of Table 4.1. The first two columns of Table 4.1 report information of married couples during the sample period. On average, wives have slightly higher BMI and less education than husbands. Roughly 25% of couples are obese or overweight in our sample. Two thirds of couples in our sample have at least one child and about half of them have health insurance. In the couple sample, 67% of men are smokers while only 3% of women smoke frequently; men have higher employment rate than women (82.3 % v.s. 70.7%). Around 30% of the couples come from urban areas. China uses a residence registration system called "hukou", which classifies people as rural or urban residents, to restrict free migration and determine eligibility to local resources such as public education, medical care and pension plan. For example, school-age children from rural areas do not have access to public schools in urban areas, even if they have been living in the urban areas. We are interested to learn whether China's economic transformation had affected overall health condition of people with urban or rural hukou differently, given that generally urban areas have benefited more from the transformation during the sample period. For this purpose, we conduct our analysis for urban and rural areas separably.

The last two columns of Tabel 4.1 report summary statistics for grown children who live in the same household with their parents. In our investigation of inter-

*The more common cut-offs are BMI between 25 to 30 for overweight and BMI greater than 30 for obese. Our analysis is not sensitive to this alternative criterion.

Table 4.1: Variables and descriptive statistics

	Married couple		Grown children	
	Female	Male	Female	Male
Non-indicator variables				
BMI	22.96	22.86	21.14	22.03
	(3.79)	(3.30)	(3.15)	(3.35)
Family income (in RMB)	30,894	30,894	29,251	35,371
	(48,970)	(48,970)	(40,180)	(64,766)
Education (in years)	6.40	8.04	9.68	9.58
	(4.32)	(3.72)	(3.60)	(3.04)
Age	43.24	45.11	23.59	27.35
	(12.17)	(12.51)	(5.44)	(7.21)
Indicator variables				
Obese (1 if BMI $\geq$ 28)	0.077	0.069	0.020	0.054
Overweight (1 if 24 $\leq$ BMI $<$ 28)	0.260	0.252	0.101	0.175
Has child	0.660	0.660	0.049	0.124
health insurance	0.479	0.516	0.534	0.573
Smoker	0.030	0.672	0.016	0.535
Employed	0.707	0.823	0.759	0.828
Urban	0.291	0.291	0.324	0.290
Married	1.000	1.000	0.132	0.492
Number of Observations	11,541	11,541	1,437	2,982

Note: Standard deviations reported in parentheses for non-indicator variables

generational BMI transmission from parents, we focus our analysis on children with complete information on both parents in the survey. In our sample there are 1,437 grown daughters and 2,982 grown sons who lived with their parents. On average, female children have lower BMI than male children. This could be partially explained by their younger age (23.59 v.s. 27.35). Unlike their parents, male and female children have the same level of average education. Only 7.5% of female children are obese or overweight, in contrast to 15.9% for male children. Not surprisingly more men are smokers than women (53.5% v.s. 1.6%); nonetheless, the prevalence of smoking is lower than that among their parents. Employment rates are 75.9% and 82.8% for female and male respectively. The employment gender gap among growth children is smaller than that of their parents (6.9% v.s. 11.6%), probably reflecting an increasing status of younger women in China's labor market. There are more married men than married women in the grown children sample (49.2% v.s. 13.2%). This is due to the fact that in China married daughters are considerably less likely to live in the same household with their parents than married sons.

4.3 Models and results

Following Abrevaya and Tang (2011), we employ proxy-variable regressions and correlated random-effects models to account for potential endogeneity and examine familial relationship of BMI. Section 4.3.1 examines impacts of demographic and economic characteristics and spousal BMI on individual BMI. Particularly we include spousal information to control for potential endogeneity due to omitted variables in the determinants of BMI. Section 4.3.2 investigates spousal BMI regressions that allow for correlated random effects to address endogeneity concern. We explore intergenerational BMI transmission between parents and grown children in Section 4.3.3.

4.3.1 Spousal BMI regressions with proxy variables

We consider the following models for wife's and husband's BMI:

$$\text{BMI}_w = x_w'\beta_w + \alpha_w + \epsilon_w \tag{4.1}$$

$$\text{BMI}_h = x_h'\beta_h + \alpha_h + \epsilon_h \tag{4.2}$$

where subscript w denotes wife and h husband. On the right hand side of both equations, x is a vector of observed individual variables, α represents an unobserved factor that may correlate with both BMI and x, and ϵ is an idiosyncratic error term satisfying $E(x'\epsilon) = 0$. We assume different coefficients in equations (4.1) and (4.2) to allow for different marginal effects of x_w and x_h on wife's and husband's BMI. We include the unobserved α to account for the fact that some likely contributing factors such as exercise or eating habits are not available from the survey and these omitted variables are possibly correlated with some covariates in x. Due to the presence of unobserved α, regressing BMI on x using the ordinary least squares (OLS) produces inconsistent results. To mitigate endogeneity bias, we employ spousal information as proxy variables for the omitted term α in the regressions. Especially we use spousal covariates x_{spouse} such as spouse's education, employment status and smoking behavior as proxies for α and incorporate them as additional covariates.

We consider three model specifications to analyze determinants of individual BMI and present the corresponding results in Tables 4.2, 4.3 and 4.4. Model 1 includes only individual characteristics as covariates with no spousal information. Model 2 considers both spousal BMI and individual characteristics to explain variations in BMI. The last model employs individual characteristics and additional spousal information as proxy for omitted variables in the regression. We include province dummies in all three models and report robust standard errors clustered at the couple level. As we discussed earlier we have four subsamples (1990s/2000s groups and urban/rural groups) for each model, and we shall use years$\times$region to denote a specific group for notational simplicity in the following discussion.

Regression results reported in Tables 4.3 and 4.4 suggest significant influence of spousal BMI on that of the other half. Husband's BMI has a larger impact on wife's BMI across all groups and models. This asymmetry is most pronounced in the 2000s$\times$urban group where husband's impact is twice as large as wife's impact, though their magnitudes have decreased over time. Overall these impacts vary over time and across regions. Comparing results across different regions from the same period, we find larger spousal impacts in the rural group after year 2000. For instance in model 2 for the 2000s group, the marginal effect of spousal's BMI in the rural sample is considerably larger than its urban counterpart (0.103 v.s. 0.064 for wife's BMI and

Table 4.2: Regression results for couple BMI in Model 1

Years & Region	Dependent variable = wife BMI				Dependent variable = husband BMI			
	1990s Urban	1990s Rural	2000s Urban	2000s Rural	1990s Urban	1990s Rural	2000s Urban	2000s Rural
Income	0.233*	0.424***	−0.061	−0.010	0.501***	0.385***	0.162*	0.127*
	(0.133)	(0.060)	(0.187)	(0.070)	(0.120)	(0.052)	(0.093)	(0.067)
Child	−0.292	−0.238*	−0.056	0.187	−0.261	−0.198	0.118	0.190
	(0.218)	(0.127)	(0.266)	(0.127)	(0.209)	(0.131)	(0.197)	(0.117)
Health insurance	−0.032	0.218*	0.092	0.435***	0.073	0.180	−0.032	0.521***
	(0.090)	(0.118)	(0.331)	(0.115)	(0.111)	(0.122)	(0.205)	(0.113)
Smoker	0.143	−0.268	−0.259	−0.180	−0.435**	−0.338***	−0.149	−0.361***
	(0.659)	(0.313)	(0.196)	(0.188)	(0.221)	(0.113)	(0.169)	(0.112)
Education	0.035	0.022	−0.088**	−0.012	0.022	0.037**	−0.021	0.101***
	(0.039)	(0.019)	(0.041)	(0.022)	(0.032)	(0.019)	(0.031)	(0.021)
Age	0.316***	0.252***	0.150	0.334***	0.219***	0.167***	0.130***	0.194***
	(0.064)	(0.034)	(0.098)	(0.036)	(0.055)	(0.031)	(0.047)	(0.033)
Age^2	−0.003***	−0.003***	−0.001	−0.003***	−0.002***	−0.002***	−0.001***	−0.002***
	(0.001)	(0.0004)	(0.001)	(0.0004)	(0.001)	(0.0004)	(0.0005)	(0.0003)
Employed	−0.485*	−0.968***	−0.669**	−0.566***	−0.347	−0.907***	0.142	−0.758***
	(0.265)	(0.189)	(0.284)	(0.136)	(0.308)	(0.210)	(0.236)	(0.170)
Observations	1,391	4,011	1,969	4,170	1,391	4,011	1,969	4,170
Adjusted R^2	0.144	0.117	0.061	0.101	0.125	0.122	0.047	0.123

Note: $*, **, ***$ denote statistical significance at the 10 %, 5% and 1% levels respectively. Robust standard errors clustered at the couple level are reported in parentheses. Province dummies are included in all models.

0.098 v.s. 0.032 for husband's BMI). Similar findings are reported in model 3 wherein more individual and spousal characteristics are included.

In models 1, 2 and 3 we investigate how individual characteristics can explain their own BMI with and without spousal BMI. Coefficients of age and age square are significantly positive and negative in all models, suggesting that BMI tends to increase with age at a decreasing rate. Smoking turns out to be a strong predictor for male BMI except for the 2000s×urban group. This is not uncommon since nicotine is known to contribute to weight loss. No significant results are observed on female BMI. This is plausibly due to lack of variation in the smoking variable for the female sample, wherein only 3% are smokers.

Income is found to be a positive predictor of husbands' BMI, and its impact seems

Table 4.3: Regression results for couple BMI in Model 2

	Dependent variable = wife BMI				Dependent variable = husband BMI			
Years & Region	1990s Urban	1990s Rural	2000s Urban	2000s Rural	1990s Urban	1990s Rural	2000s Urban	2000s Rural
Spousal BMI	0.155***	0.151***	0.064*	0.103***	0.141***	0.117***	0.032*	0.098***
	(0.036)	(0.025)	(0.034)	(0.025)	(0.033)	(0.020)	(0.017)	(0.022)
Income	0.157	0.371***	−0.074	−0.024	0.471***	0.334***	0.166*	0.129*
	(0.133)	(0.060)	(0.186)	(0.070)	(0.121)	(0.052)	(0.093)	(0.066)
Child	−0.249	−0.207	−0.062	0.171	−0.219	−0.170	0.118	0.171
	(0.214)	(0.128)	(0.265)	(0.127)	(0.204)	(0.131)	(0.196)	(0.116)
Health insurance	−0.018	0.201*	0.107	0.384***	0.078	0.146	−0.043	0.480***
	(0.091)	(0.110)	(0.326)	(0.114)	(0.111)	(0.114)	(0.204)	(0.111)
Smoker	0.219	−0.221	−0.240	−0.146	−0.432*	−0.336***	−0.144	−0.364***
	(0.639)	(0.313)	(0.194)	(0.191)	(0.221)	(0.113)	(0.169)	(0.112)
Education	0.029	0.011	−0.088**	−0.019	0.015	0.036**	−0.018	0.101***
	(0.038)	(0.019)	(0.041)	(0.022)	(0.031)	(0.018)	(0.031)	(0.021)
Age	0.285***	0.227***	0.141	0.317***	0.170***	0.140***	0.129***	0.165***
	(0.064)	(0.034)	(0.097)	(0.037)	(0.056)	(0.031)	(0.047)	(0.034)
Age^2	−0.002***	−0.002***	−0.001	−0.003***	−0.002***	−0.001***	−0.001***	−0.002***
	(0.001)	(0.0004)	(0.001)	(0.0004)	(0.001)	(0.0004)	(0.0005)	(0.0004)
Employed	−0.394	−0.885***	−0.649**	−0.518***	−0.298	−0.785***	0.133	−0.728***
	(0.260)	(0.187)	(0.281)	(0.135)	(0.299)	(0.209)	(0.236)	(0.168)
Observations	1,391	4,011	1,969	4,170	1,391	4,011	1,969	4,170
Adjusted R^2	0.162	0.132	0.063	0.110	0.143	0.137	0.049	0.132

Note: $*, **, ***$ denote statistical significance at the 10 %, 5% and 1% levels respectively. Robust standard errors clustered at the couple level are reported in parentheses. Province dummies are included in all models.

Table 4.4: Regression results for couple BMI in Model 3

Years & Region	Dependent variable = wife BMI				Dependent variable = husband BMI			
	1990s Urban	1990s Rural	2000s Urban	2000s Rural	1990s Urban	1990s Rural	2000s Urban	2000s Rural
Spousal BMI	0.154***	0.149***	0.062*	0.104***	0.137***	0.112***	0.030*	0.096***
	(0.036)	(0.025)	(0.034)	(0.025)	(0.032)	(0.020)	(0.017)	(0.023)
Spouse education	0.039	−0.015	−0.064*	−0.017	0.010	0.070***	0.005	0.025
	(0.039)	(0.023)	(0.037)	(0.026)	(0.035)	(0.018)	(0.034)	(0.023)
Spouse employed	−0.124	−0.576*	0.239	−0.095	−0.580**	−0.201	−0.352*	−0.241*
	(0.321)	(0.299)	(0.267)	(0.191)	(0.239)	(0.171)	(0.195)	(0.135)
Income	0.142	0.377***	−0.060	−0.014	0.476***	0.292***	0.180*	0.125*
	(0.135)	(0.061)	(0.183)	(0.071)	(0.123)	(0.052)	(0.094)	(0.066)
Child	−0.246	−0.214*	−0.060	0.162	−0.216	−0.162	0.097	0.171
	(0.213)	(0.128)	(0.265)	(0.126)	(0.203)	(0.130)	(0.198)	(0.117)
Health insurance	−0.025	0.190*	0.109	0.387***	0.085	0.109	−0.038	0.472***
	(0.091)	(0.109)	(0.327)	(0.114)	(0.113)	(0.107)	(0.206)	(0.111)
Smoker & non-smoker	1.060	0.282	−0.112*	0.134	−0.414*	−0.319***	−0.148	−0.366***
	(0.936)	(0.602)	(0.065)	(0.262)	(0.227)	(0.113)	(0.170)	(0.112)
Smoker & smoker	−0.182	−0.300	−0.923	−0.231	−1.042*	−0.687***	−0.641	−0.700***
	(0.730)	(0.361)	(0.819)	(0.222)	(0.555)	(0.266)	(0.893)	(0.183)
Non-smoker & smoker	0.078	0.064	−0.213	0.056	−0.200	0.169	−0.233***	−0.322
	(0.218)	(0.124)	(0.239)	(0.108)	(0.659)	(0.834)	(0.044)	(0.292)
Education	0.009	0.015	−0.063	−0.015	0.009	0.011	−0.017	0.089***
	(0.042)	(0.021)	(0.043)	(0.025)	(0.032)	(0.019)	(0.035)	(0.023)
Age	0.291***	0.237***	0.138	0.318***	0.175***	0.156***	0.127***	0.174***
	(0.064)	(0.035)	(0.101)	(0.037)	(0.056)	(0.032)	(0.047)	(0.034)
Age^2	−0.002***	−0.003***	−0.001	−0.003***	−0.002***	−0.002***	−0.001***	−0.002***
	(0.001)	(0.0004)	(0.001)	(0.0004)	(0.001)	(0.0004)	(0.0005)	(0.0004)
Employed	−0.358	−0.681***	−0.688**	−0.492***	−0.075	−0.623***	0.209	−0.600***
	(0.263)	(0.213)	(0.285)	(0.141)	(0.313)	(0.235)	(0.233)	(0.180)
Observations	1,391	4,011	1,969	4,170	1,391	4,011	1,969	4,170
Adjusted R^2	0.162	0.134	0.062	0.109	0.146	0.144	0.049	0.134

Note: $*, **, ***$ denote statistical significance at the 10 %, 5% and 1% levels respectively. Robust standard errors clustered at the couple level are reported in parentheses. The second entry in smoking categories indicates if the spouse is a smoker, and the omitted smoking category in model 4 is both non-smokers. Province dummies are included in all models.

to decline over time. Since nutrition demand is higher for men than women, it is not unexpected to see a positive impact of increased income on male BMI during early stage of economic development. As China went through further economic transformation, nutritious food became more readily available. Correspondingly, we observe in our analysis diminished income effect for male BMI in the 2000s. For women, significant income effect is only observed for the 1990s×urban group. We note that income effect on male BMI also varies regions. For example in model 1 coefficients of income are 0.501 and 0.385 for male from the urban and rural areas in the 1990s; they are reduced to 0.162 and 0.127 in the 2000s. Similar findings appear in models 2 and 3.

Employment status is a useful predictor of BMI for some groups. Men in the rural area are more likely to have lower BMI if they are employed, while this employment effect is not significant in the urban area, probably reflecting the fact that jobs for men in the country tend to be more labor intensive than those in the cities. Women in all but the 1990s×urban group are also likely to have lower BMI if they are employed.

What if spousal characteristics, other than BMI, are used as proxy for omitted determinants of individual BMI? To answer this question we include additional spousal information such as education, employment status and smoking in model 3. We also include an interaction term between couples' smoking indicators, treating the category of non-smoking husband and wife as the baseline in the model. Spousal education is a significant predictor for female BMI (-0.064) in the 2000s×urban group and for male BMI (0.070) in the 1990s×rural group. Interestingly though the magnitudes are close they show opposite signs, suggesting that spousal education probably proxies for different latent components in the omitted term α. Spouse's employment status is also a useful predictor for male BMI except for the 1990s×rural group. In particular, it is suggested that the husband tends to have lower BMI if his wife is employed. One plausible reason for this finding is that in Chinese households, wives are typically responsible for most of house keeping and meal preparation. If the wife is employed, she would have less time for meal preparation and at the same time the husband might have an increased share of house keeping, both of which might have a negative effect on husband's BMI.

Spouse's smoking seems to be a complement of own smoking for men, as the coefficients of double-smokers are at least 100% higher than those of smoking husband and non-smoking wife. We conjecture that this interaction term proxies for some unobserved factors such as health awareness or living environment. These terms are significant for female BMI only in one group (2000s×urban). Again this is probably due to the fact that only 3% of wives are smokers in our sample. Overall we find that incorporating spousal information as proxy variables tends to improve the prediction of husband's BMI, as is evident from higher adjusted R-squares in those regressions.

4.3.2 Spousal BMI regressions with correlated random effects

In this subsection we employ the correlated random-effcts (CRE) model (Chamberlain 1982) to account for potential dependence between unobserved common factors of spousal BMI and observable individual characteristics x. Specifically we consider the following models for wife's and husband's BMI:

(4.3) $\mathrm{BMI}_w = x_w \beta_w + \alpha + \epsilon_w$

(4.4) $\mathrm{BMI}_h = x_h \beta_h + \alpha + \epsilon_h$

where $\alpha = \alpha_w = \alpha_h$ is a common component for wife and husband, which we shall term the "couple effect". Chamberlain (1982) treats α as a linear projection onto the observed regressors x such that:

(4.5) $\alpha = \phi + x_w\lambda_w + x_h\lambda_h + \upsilon$

where ϕ is the intercept and υ is an error term uncorrelated with x_w and x_h by construction of linear projection. An important advantage of the CRE estimator is that λ_w and λ_h directly show which of the observable variables x are correlated with the unobserved common component α. Combining equations (4.3) and (4.4) with projection in (4.5), we obtain

(4.6) $\text{BMI}_w = \phi + x_w\beta_w + x_h\lambda_h + x_w\lambda_w + (\upsilon + \epsilon_w)$

(4.7) $\text{BMI}_h = \phi + x_h\beta_h + x_h\lambda_h + x_w\lambda_w + (\upsilon + \epsilon_h)$

We estimate equations (4.6) and (4.7) using the pooled OLS regression as suggested by Wooldridge (2010). Notice that CRE model cannot identify couple effects λ for variables that are shared by husband and wife (income and child indicator in our case). Therefore we should interpret the coefficients β of these shared variables as the overall effects that are measured by $\beta+\lambda$ for non-shared variables. Estimation results of CRE models are reported in Tables 4.5 and 4.6 for samples in different years and areas. Specifically Table 4.5 presents results for 1990s sample and Table 4.6 results for 2000s sample. For comparison between OLS model and CRE model, below we shall frequently refer back to the results of couple model 1 in Table 4.2 from Section 4.3.1.

Results of the CRE model for the 1990s groups are similar to those in couple model 1. The coefficients for β_{smoker} in the husband equation are negative and their magnitudes are close to those in couple model 1 (-0.435 v.s. -0.435 for urban area, -0.306 v.s. -0.338 for rural area), while their couple effects λ_{smoker} are not

Table 4.5: CRE model BMI regression results for couples in urban and rural area during 1990s

Models	1990s urban couples				1990s rural couples			
	β estimates		λ estimates		β estimates		λ estimates	
Variables	Husband	Wife	Husband	Wife	Husband	Wife	Husband	Wife
Male	1.862				1.667*			
	(1.905)				(0.973)			
Income	0.461***	0.248**			0.328***	0.440***		
	(0.101)	(0.112)			(0.047)	(0.053)		
Child	−0.313*	−0.268			−0.199*	−0.246**		
	(0.182)	(0.188)			(0.108)	(0.110)		
Health insurance	0.216	0.204	0.007	−0.235**	−0.166	−0.045	0.362**	−0.075
	(0.166)	(0.145)	(0.109)	(0.106)	(0.236)	(0.230)	(0.177)	(0.151)
Smoker	−0.435*	0.608	−0.005	−0.422	−0.306**	0.066	−0.021	−0.338
	(0.247)	(0.720)	(0.177)	(0.423)	(0.134)	(0.352)	(0.099)	(0.227)
Education	−0.040	−0.006	0.048*	0.021	0.017	−0.048**	−0.005	0.075***
	(0.038)	(0.040)	(0.028)	(0.027)	(0.022)	(0.019)	(0.017)	(0.013)
Age	−0.194	−0.091	0.374***	0.054	0.017	0.090	0.146**	0.027
	(0.131)	(0.139)	(0.100)	(0.093)	(0.080)	(0.084)	(0.059)	(0.057)
Age^2	0.001	0.0004	−0.003***	0.00005	−0.001	−0.002*	−0.001	−0.0003
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Employed	0.115	0.193	−0.179	−0.597***	−0.087	−0.406	−0.620**	−0.289*
	(0.430)	(0.312)	(0.312)	(0.217)	(0.346)	(0.250)	(0.267)	(0.157)

Note: $*, **, ***$ denote statistical significance at the 10 %, 5% and 1% levels respectively. Robust standard errors are in parentheses. Province dummies are included in all regressions.

statistically significant. This result implies that smoking does not have an indirect impact on husband's BMI through the shared couple effect. CRE models also suggest no significant direct effect (β_{smoker}) or indirect effect (λ_{smoker}) on wife's BMI for the 1990s group, which is consistent with the findings in couple model 1. Interestingly we find evidence of couple effects for women in the 2000s group, where λ_{smoker} are -0.326 and -0.321 for the urban and rural areas respectively. Both coefficients are significant at least at the 10% level, while neither β_{smoker} in couple model 1 nor CRE model is statistically significant. This discrepancy suggests that ignoring common couple effects might lead to an upward bias in the estimate of smoking effect on women's BMI, as a negative λ_{smoker} implies there may exist omitted variables in the couple effect α that are negatively correlated with smoking. Therefore although

Table 4.6: CRE model BMI regression results for couples in urban and rural area during 2000s

Models	2000s urban couples				2000s rural couples			
	β estimates		λ estimates		β estimates		λ estimates	
Variables	Husband	Wife	Husband	Wife	Husband	Wife	Husband	Wife
Male	−0.674				0.900			
	(3.793)				(1.179)			
Income	0.150*	−0.011			0.133**	−0.021		
	(0.084)	(0.157)			(0.060)	(0.060)		
Child	−0.025	0.073			0.201**	0.165		
	(0.181)	(0.218)			(0.100)	(0.106)		
Health insurance	−0.340	0.341	0.458	−0.380	0.046	−0.035	0.198	0.304
	(0.500)	(0.635)	(0.447)	(0.237)	(0.309)	(0.303)	(0.225)	(0.207)
Smoker	0.014	0.088	−0.188	−0.326*	−0.410***	0.110	0.028	−0.321***
	(0.274)	(0.257)	(0.228)	(0.181)	(0.124)	(0.195)	(0.086)	(0.118)
Education	0.027	−0.067	−0.052*	−0.0001	0.094***	−0.049**	−0.004	0.033*
	(0.041)	(0.052)	(0.030)	(0.028)	(0.026)	(0.025)	(0.020)	(0.017)
Age	0.576	0.568*	−0.501	0.058	0.269***	0.395***	−0.089	0.019
	(0.427)	(0.332)	(0.419)	(0.091)	(0.100)	(0.096)	(0.073)	(0.065)
Age^2	−0.007	−0.006	0.006	−0.0004	−0.004***	−0.005***	0.001*	0.0003
	(0.005)	(0.004)	(0.005)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Employed	−0.100	−0.466	0.310	−0.325*	−0.532**	−0.161	−0.086	−0.323***
	(0.349)	(0.310)	(0.293)	(0.183)	(0.231)	(0.173)	(0.167)	(0.120)

Note: $*, **, ***$ denote statistical significance at the 10 %, 5% and 1% levels respectively. Robust standard errors are in parentheses. Province dummies are included in all regressions.

smoking does not have a direct impact on women's BMI in our analysis, it can exert influence through couple effects via living environment, health awareness, exercise and eating habits, etc.

Education is of great interest in our study since it is potentially correlated with many unobservables in the common effect α. Before we proceed to the discussion of education effect we want to emphasize that income effects in couple model 1 and CRE model are close in magnitude in all four subsamples, therefore we have similar control for income in both models and the following analysis of education impacts on BMI should work through channels other than income. We first discuss the results for men in the urban area. For the 1990s group, β_{edu} is insignificant in couple model 1 and CRE model, but the couple effect λ_{edu}, estimated at 0.048, is marginally significant at the 10% level. For the 2000s group, education again exhibits little direct impact on BMI, while λ_{edu} is estimated at -0.052 with a 10% significance level. It appears that education of men living in the cities is correlated with couple effect α in a time-varying manner. One plausible explanation for this apparent sign switch in education impacts is that better educated people in the cities are generally more health conscious and tend to adopt healthy diet and life style; the benefits of healthy habits are initially manifested by higher BMI in the early stage of economic development, followed by lower BMI under a more advanced economy. We also note that after controlling for couple effect men in the 1990s×rural group do not have a significant overall education effect ($\beta_{\text{edu}}+\lambda_{\text{edu}}$) in CRE model, while β_{edu} is 0.037 and significant at the 5% level in couple model 1. This difference suggests that estimate of couple model 1 might be biased upward due to omitted variables.

Couple effects through education are also found for women, which appear to decrease in the rural area over time. For the 1990s×rural group, β_{edu} of couple model 1 is insignificant, but in the CRE model $\beta_{\text{edu}} = -0.048$ and $\lambda_{\text{edu}} = 0.075$, both of which are significant at least at the 5% level. Results for the 2000s×rural group are similar ($\beta_{\text{edu}} = -0.049$ and $\lambda_{\text{edu}} = 0.033$). Couple effect λ_{edu} decreases by 56% from 1990s to 2000s though it remains statistically significant. The evolvement of λ_{edu} suggests that the association between education and omitted variables has changed over time. For the women in 2000s×urban group, β_{edu} of couple model 1 is -0.088 and significant at the 5% level, but the CRE model shows that both β_{edu} and λ_{edu} are not significant, implying that β_{edu} is biased downward when couple effect is not taken into account.

4.3.3 BMI regressions for grown children

In this section we analyze grown children's BMI. We are especially interested in intergenerational transmission of BMI from parents to children. Compared to the parents sample studied in Section 4.3.1 and 4.3.2, grown children in our sample are on the average better educated and more likely to be employed and have insurance. We examine daughter's BMI and son's BMI separably to allow for gender-specific impacts of various contributing factors. With parental characteristics as proxies to account for omitted variable issues, our estimation strategy is similar to those in Section 4.3.1. Particularly we consider two models: (1) linear regression of children BMI on parental BMI and individual characteristics. (2) incorporating additional parental characteristics as proxy variables. We report estimation results for these models in Tables 4.7 and 4.8 respectively.

Our results suggest that parental BMI is a strong predictor of children BMI. Father's BMI in general has a greater impact except for children in the 1990s$\times$urban group. We note the parental BMI impact depends on the gender of children and varies across areas. For daughters living in the urban area BMI effects of mother and father have increased over time (from 0.089 to 0.154 and from 0.066 to 0.400 respectively) in model 2. At the same time these effects have declined for daughters in the rural area. For sons living either in urban or rural area, the BMI impact of father is highly significant (at 1% level) and has increased over time (from 0.161 to 0.225 for the urban group and from 0.263 to 0.404 for the rural group) in model 2, while the BMI impact of mother has increased only for sons in the rural group. Similar patterns are observed in model 1. The overall results indicate that genetic traits shared by parents and children play a critical role in intergenerational BMI transmission.

Table 4.7: Regression results for growth children BMI in Model 1

Years & Region	Dependent variable = daughter BMI				Dependent variable = son BMI			
	1990s Urban	1990s Rural	2000s Urban	2000s Rural	1990s Urban	1990s Rural	2000s Urban	2000s Rural
Mother BMI	0.069	0.121***	0.150*	0.101*	0.170***	0.149***	0.133***	0.150***
	(0.043)	(0.031)	(0.080)	(0.059)	(0.034)	(0.018)	(0.045)	(0.035)
Father BMI	0.084**	0.163***	0.399***	0.104**	0.152***	0.268***	0.224***	0.390***
	(0.039)	(0.036)	(0.134)	(0.049)	(0.038)	(0.032)	(0.041)	(0.054)
Income	0.222	0.087	0.554*	0.417**	0.281	−0.104	0.251	0.035
	(0.183)	(0.116)	(0.337)	(0.205)	(0.173)	(0.072)	(0.198)	(0.127)
Child	0.692	−0.139	−1.071	−0.051	−0.333	0.073	−0.026	−0.339
	(1.000)	(0.740)	(0.700)	(0.856)	(0.506)	(0.272)	(0.472)	(0.266)
Health insurance	0.516***	−0.568**	0.363	−0.054	0.212**	0.382**	0.313	0.225
	(0.137)	(0.280)	(0.552)	(0.379)	(0.107)	(0.163)	(0.363)	(0.223)
Smoker	−0.627	1.336**	−0.549	−1.617**	0.084	0.070	−0.014	−0.795***
	(1.283)	(0.641)	(1.578)	(0.639)	(0.256)	(0.129)	(0.365)	(0.229)
Education	−0.224***	−0.145***	−0.149**	−0.112	−0.058	0.050**	0.032	0.048
	(0.055)	(0.031)	(0.074)	(0.084)	(0.052)	(0.024)	(0.067)	(0.036)
Age	0.062	−0.005	−0.247	−0.196	0.183*	0.225**	0.551***	0.515***
	(0.181)	(0.136)	(0.309)	(0.222)	(0.099)	(0.099)	(0.130)	(0.092)
Age^2	0.0004	0.001	0.005	0.006	−0.001	−0.002	−0.007***	−0.006***
	(0.003)	(0.002)	(0.005)	(0.005)	(0.001)	(0.002)	(0.002)	(0.001)
Employed	−0.632	0.142	−0.187	−0.229	0.779**	0.237	0.319	−0.049
	(0.394)	(0.285)	(0.585)	(0.339)	(0.350)	(0.245)	(0.452)	(0.284)
Married	−0.293	0.604	1.273**	0.333	−0.136	0.241	0.645	0.118
	(0.705)	(0.557)	(0.634)	(0.799)	(0.341)	(0.169)	(0.449)	(0.282)
Observations	240	647	225	325	429	1,128	437	988
Adjusted R^2	0.098	0.108	0.091	0.069	0.223	0.239	0.175	0.223

Note: $*, **, ***$ denote statistical significance at the 10 %, 5% and 1% levels respectively. Robust standard errors are in parentheses. Province dummies are included in all regressions.

Table 4.8: Regression results for growth children BMI in Model 2

Years & Region	Dependent variable = daughter BMI				Dependent variable = son BMI			
	1990s Urban	1990s Rural	2000s Urban	2000s Rural	1990s Urban	1990s Rural	2000s Urban	2000s Rural
Mother BMI	0.089**	0.131***	0.154*	0.110*	0.178***	0.141***	0.133***	0.153***
	(0.044)	(0.032)	(0.080)	(0.059)	(0.034)	(0.019)	(0.046)	(0.034)
Mother smoker	−0.406	0.255	0.358	−0.386	−0.165	−0.140	0.576	−0.514
	(0.559)	(0.463)	(2.385)	(0.836)	(0.480)	(0.368)	(0.862)	(0.480)
Mother education	−0.090**	−0.022	0.010	0.096**	−0.004	−0.021	−0.033	0.040
	(0.044)	(0.028)	(0.078)	(0.049)	(0.040)	(0.026)	(0.045)	(0.033)
Mother employed	0.396	0.201	−0.737	0.223	−0.081	−0.289	−1.233**	−0.022
	(0.328)	(0.284)	(0.872)	(0.355)	(0.258)	(0.208)	(0.479)	(0.260)
Father BMI	0.066*	0.177***	0.400***	0.102**	0.161***	0.263***	0.225***	0.404***
	(0.039)	(0.036)	(0.136)	(0.047)	(0.038)	(0.032)	(0.039)	(0.054)
Father smoker	−0.549*	0.216	−0.007	−0.029	0.121	−0.090	−0.012	0.482**
	(0.330)	(0.209)	(0.341)	(0.173)	(0.252)	(0.152)	(0.298)	(0.212)
Father education	0.008	−0.090***	−0.061	−0.031	−0.067*	−0.003	0.008	−0.058
	(0.038)	(0.028)	(0.103)	(0.058)	(0.034)	(0.022)	(0.044)	(0.040)
Father employed	−0.096	0.282	1.086	−0.281	−0.283	−0.313	0.312	0.345
	(0.328)	(0.464)	(0.701)	(0.425)	(0.286)	(0.281)	(0.405)	(0.283)
Income	0.223	0.119	0.540	0.411**	0.350**	−0.095	0.290	0.003
	(0.190)	(0.116)	(0.410)	(0.200)	(0.170)	(0.073)	(0.205)	(0.126)
Child	0.817	−0.208	−1.123	−0.027	−0.374	0.087	−0.043	−0.401
	(0.958)	(0.754)	(0.699)	(0.858)	(0.510)	(0.271)	(0.472)	(0.265)
Health insurance	0.507***	−0.557**	0.520	−0.103	0.194*	0.332**	0.368	0.238
	(0.129)	(0.279)	(0.603)	(0.385)	(0.107)	(0.163)	(0.364)	(0.222)
Smoker	−0.417	1.307**	−0.046	−1.115*	0.025	0.086	−0.086	−0.797***
	(1.360)	(0.652)	(1.484)	(0.665)	(0.256)	(0.130)	(0.372)	(0.228)
Education	−0.153**	−0.101***	−0.147*	−0.143	−0.011	0.057**	0.032	0.058
	(0.062)	(0.035)	(0.087)	(0.089)	(0.055)	(0.023)	(0.066)	(0.036)
Age	0.086	−0.015	−0.230	−0.200	0.116	0.213**	0.473***	0.544***
	(0.175)	(0.133)	(0.283)	(0.225)	(0.102)	(0.101)	(0.133)	(0.092)
Age^2	−0.0004	0.002	0.005	0.006	−0.0002	−0.002	−0.007***	−0.007***
	(0.003)	(0.002)	(0.004)	(0.005)	(0.001)	(0.002)	(0.002)	(0.001)
Employed	−0.550	−0.028	−0.162	−0.151	0.782**	0.373	0.207	−0.159
	(0.381)	(0.296)	(0.637)	(0.345)	(0.346)	(0.248)	(0.448)	(0.296)
Married	−0.449	0.597	1.416**	0.400	−0.163	0.216	0.606	0.201
	(0.714)	(0.549)	(0.658)	(0.786)	(0.344)	(0.168)	(0.434)	(0.287)
Observations	240	647	225	325	429	1,128	437	988
Adjusted R^2	0.106	0.124	0.076	0.062	0.226	0.241	0.178	0.227

Note: $*, **, ***$ denote statistical significance at the 10 %, 5% and 1% levels respectively. Robust standard errors are in parentheses. Province dummies are included in all regressions.

We find negative and significant education effect for daughters' BMI except for those in the 2000s×rural group. This negative effect is stronger for the urban groups in both models. Notice that in section 4.3.1 we do not have similar results from spousal BMI models, even though their sample sizes are 6 to 10 times larger than the children sample. Considering that income and employment status have been controlled in our models, this result suggests that young educated women in the cities probably are more health conscious and more likely to control their weights via healthy diet and life style. Positive education effect is also found for sons in the 1990s×rural group, significant at the 5% level.

Comparing model 1 and 2 suggests that incorporating additional parental information as proxy variables does not necessarily produce better models. For example, adjusted R-squares in model 1 are only slightly smaller than those in model 2 for daughters in the 1990s group, while for those in the 2000s group these values are higher in model 1 where only parental BMI is used as proxy variable. Similar patterns are observed for sons' regressions. Recall that in the couples' models above, incorporating additional spousal information generally improves the performance. We conjecture that the lack of improvement in the regressions on children's BMI is because the common living environment, proxies by spousal characteristics, is the main reason behind the dependence in a couple's BMI; in contrast, generic linkage is the vastly dominant factor in intergenerational BMI transmission.

4.4 Conclusion

This paper uses a nationally representative household survey of China to study familial relationship in BMI between 1991 through 2011. We find positive effects of spousal BMI that are significant, asymmetric (greater for wife than for husband) and generally vary across regions and over time. Income is found to be a strong positive predictor for husband's BMI, while employment has a significant and negative impact on wife's BMI. Similar to Abrevaya and Tang (2011) we find significant couple effect shared by wife and husband for education level in the correlated random-effects models. For grown children, we find parental BMI to be the most important predictors for children's BMI. Education attainment is shown to have a negative impact on daughters' BMI. Since families can play an essential role in preventing obesity, our results can be useful for developing health intervention programs and promoting healthy lifestyle.

5. CONCLUSION

In the first essay I propose to model time-varying dependence structure of commodity markets with dynamic copulas. This advanced method enables dynamic modeling for high-dimensional data. We show that during the 2008-2009 financial crisis co-movement of commodity futures returns is evident and the conditional diversification benefit of commodity portfolio has decreased dramatically. After 2014 markets behavior returned to normal and move exactly like what they did before the crisis.

The second essay focuses on network connectedness of commodity markets with VAR models. My static and dynamic analysis show that the connectedness measures for commodity markets are strikingly different during various economic and financial cycles. My results suggest clustering of commodity markets which match their industrial groupings. I also show that energy markets have played a major role in the commodity markets until 2013 and resumed its status since the beginning of 2015. This essay as well as the first essay, however, do not fully take macroeconomic variables into account. One obvious reason is that macro data is usually available in a low frequency manner I suggest futures research may consider how to combine variables of different frequencies to shed light on effects of macroeconomy on co-movements in commodity markets.

My last essay specializes in obesity and overweight problem in China. Using a large micro-data over twenty years across urban and rural regions, we find multiple factors that can contribute to individual BMI. I find the transmission effects of spousal and intergenerational BMI are significant and robust using various model specifications. Future research can extend the empirical results in this essay by designing more informative survey to obtain instrumental variables to handle endogeneity concern, which is largely due to the omitted variables that are not easily collected or observed in the current micro-data.

REFERENCES

Abrevaya, J., and H. Tang. 2011. "Body mass index in families: spousal correlation, endogeneity, and intergenerational transmission." *Empirical Economics* 41:841–864.

Adams, Z., and T. Glück. 2015. "Financialization in commodity markets: A passing trend or the new normal?" *Journal of Banking & Finance* 60:93–111.

Aielli, G.P. 2013. "Dynamic conditional correlation: on properties and estimation." *Journal of Business & Economic Statistics* 31:282–299.

Anderson, P.M., K.F. Butcher, and P.B. Levine. 2003. "Maternal employment and overweight children." *Journal of Health Economics* 22:477–504.

Artzner, P., F. Delbaen, J.M. Eber, and D. Heath. 1999. "Coherent measures of risk." *Mathematical Finance* 9:203–228.

Belousova, J., and G. Dorfleitner. 2012. "On the diversification benefits of commodities from the perspective of euro investors." *Journal of Banking & Finance* 36:2455–2472.

Bessler, W., and D. Wolff. 2015. "Do commodities add value in multi-asset portfolios? An out-of-sample analysis for different investment strategies." *Journal of Banking & Finance* 60:1–20.

Büyükşahin, B., and M.A. Robe. 2017. "The financialization of food?" *American Journal of Agricultural Economics* 99:243–264.

—. 2014. "Speculators, commodities and cross-market linkages." *Journal of International Money and Finance* 42:38–70.

Chamberlain, G. 1982. "Multivariate regression models for panel data." *Journal of Econometrics* 18:5–46.

Chen, H.J., Y. Liu, and Y. Wang. 2014. "Socioeconomic and demographic factors for spousal resemblance in obesity status and habitual physical activity in the United States." *Journal of Obesity* 2014:1–11.

Chou, S.Y., M. Grossman, and H. Saffer. 2004. "An economic analysis of adult obesity: results from the Behavioral Risk Factor Surveillance System." *Journal of Health Economics* 23:565–587.

Christoffersen, P., V. Errunza, K. Jacobs, and X. Jin. 2014. "Correlation dynamics and international diversification benefits." *International Journal of Forecasting* 30:807–824.

Christoffersen, P., V. Errunza, K. Jacobs, and H. Langlois. 2012. "Is the potential for international diversification disappearing? A dynamic copula approach." *Review of Financial Studies* 25:3711–3751.

Christoffersen, P., K. Jacobs, X. Jin, and H. Langlois. 2016. "Dynamic dependence and diversification in corporate credit." Unpublished, Working paper, Bauer College of Business, University of Houston.

Christoffersen, P., A. Lunde, and K.V. Olesen. 2014. "Factor structure in commodity futures return and volatility." Unpublished, Working paper, University of Toronto.

Classen, T., and C. Hokayem. 2005. "Childhood influences on youth obesity." *Economics & Human Biology* 3:165–187.

Cohen, A.K., M. Rai, D.H. Rehkopf, and B. Abrams. 2013. "Educational attainment and obesity: a systematic review." *Obesity Reviews* 14:989–1005.

Creal, D., S.J. Koopman, and A. Lucas. 2013. "Generalized autoregressive score models with applications." *Journal of Applied Econometrics* 28:777–795.

Creal, D.D., and R.S. Tsay. 2015. "High dimensional dynamic stochastic copula models." *Journal of Econometrics* 189:335–345.

Daskalaki, C., and G. Skiadopoulos. 2011. "Should investors include commodities in their portfolios after all? New evidence." *Journal of Banking & Finance* 35:2606–2626.

Demarta, S., and A.J. McNeil. 2005. "The t copula and related copulas." *International Statistical Review/Revue Internationale de Statistique*, pp. 111–129.

Demirer, M., F.X. Diebold, L. Liu, and K. Yilmaz. 2017. "Estimating global bank network connectedness." Unpublished, Working paper, University of Pennsylvania.

Diebold, F.X. 2015. "Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests." *Journal of Business & Economic Statistics* 33:1–1.

Diebold, F.X., and R.S. Mariano. 1995. "Comparing predictive accuracy." *Journal of Business & Economic Statistics* 13:134–144.

Diebold, F.X., and K. Yilmaz. 2012. "Better to give than to receive: Predictive directional measurement of volatility spillovers." *International Journal of Forecasting* 28:57–66.

—. 2015a. *Financial and macroeconomic connectedness: a network approach to measurement and monitoring*. Oxford University Press, USA.

—. 2014. "On the network topology of variance decompositions: Measuring the connectedness of financial firms." *Journal of Econometrics* 182:119–134.

—. 2015b. "Trans-Atlantic equity volatility connectedness: US and European financial institutions, 2004–2014." *Journal of Financial Econometrics* 14:81–127.

Du, X., C.L. Yu, and D.J. Hayes. 2011. "Speculation and volatility spillover in the crude oil and agricultural commodity markets: A Bayesian analysis." *Energy Economics* 33:497 – 503.

Duffie, D., J. Pan, and K. Singleton. 2000. "Transform analysis and asset pricing for affine jump-diffusions." *Econometrica* 68:1343–1376.

Engle, R. 2002. "Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models." *Journal of Business & Economic Statistics* 20:339–350.

Engle, R., and B. Kelly. 2012. "Dynamic equicorrelation." *Journal of Business & Economic Statistics* 30:212–228.

Engle, R.F., N. Shephard, and K. Sheppard. 2008. "Fitting vast dimensional time-varying covariance models." Unpublished, Working Paper, New York University.

Eraker, B., M. Johannes, and N. Polson. 2003. "The impact of jumps in volatility and returns." *The Journal of Finance* 58:1269–1300.

Gao, P., and J. Shen. 2016. "An empirical analysis on the determinants of overweight and obesity in China." *Applied Economics*, pp. 1–14.

Gao, Y., Y.Y. Zhang, and X. Wu. 2015. "Penalized exponential series estimation of copula densities with an application to intergenerational dependence of body mass index." *Empirical Economics* 48:61–81.

Garman, M.B., and M.J. Klass. 1980. "On the estimation of security price volatilities from historical data." *Journal of Business*, pp. 67–78.

Ghysels, E., P. Santa-Clara, and R. Valkanov. 2004. "The MIDAS touch: Mixed data sampling regression models." Unpublished, Working paper, UNC and UCLA.

Giacomini, R., and H. White. 2006. "Tests of conditional predictive ability." *Econometrica* 74:1545–1578.

Glasserman, P., and H.P. Young. 2016. "Contagion in financial networks." *Journal of Economic Literature* 54:779–831.

Hafner, C.M., and H. Herwartz. 2006. "A Lagrange multiplier test for causality in variance." *Economics Letters* 93:137–141.

Hansen, B.E. 1994. "Autoregressive conditional density estimation." *International Economic Review*, pp. 705–730.

Hansen, P.R., and A. Timmermann. 2012. "Choice of sample split in out-of-sample forecast evaluation.", pp. .

Inoue, A., L. Jin, and B. Rossi. 2017. "Rolling window selection for out-of-sample forecasting with time-varying parameters." *Journal of Econometrics* 196:55 – 67.

Jacomy, M., T. Venturini, S. Heymann, and M. Bastian. 2014. "ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software." *PloS one* 9:e98679.

Jeffery, R.W., and A.M. Rick. 2002. "Cross-sectional and longitudinal associations between body mass index and marriage-related factors." *Obesity Research* 10:809–815.

Kilian, L. 2009. "Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market." *American Economic Review* 99:1053–69.

Koop, G., M.H. Pesaran, and S.M. Potter. 1996. "Impulse response analysis in nonlinear multivariate models." *Journal of Econometrics* 74:119–147.

Lakdawalla, D., and T. Philipson. 2009. "The growth of obesity and technological change." *Economics & Human Biology* 7:283–293.

Lucas, A., B. Schwaab, and X. Zhang. 2017. "Modeling financial sector joint tail risk in the euro area." *Journal of Applied Econometrics* 32:171–191.

Mamun, A., D. Lawlor, M. O'Callaghan, G. Williams, and J. Najman. 2005. "Family and early life factors associated with changes in overweight status between ages 5 and 14 years: findings from the Mater University Study of Pregnancy and its outcomes." *International Journal of Obesity* 29:475–482.

National Health and Family Planning Commission of the Peoples Republic of China. 2014. *Report on Chinese residents chronic disease and nutrition*. http://en.nhfpc.gov.cn/2015-06/15/c_45788.htm.

Nazlioglu, S., C. Erdem, and U. Soytas. 2013. "Volatility spillover between oil and agricultural commodity markets." *Energy Economics* 36:658 – 665.

Ng, M., T. Fleming, M. Robinson, B. Thomson, N. Graetz, C. Margono, E.C. Mullany, S. Biryukov, C. Abbafati, S.F. Abera, et al. 2014. "Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013." *The Lancet* 384:766–781.

Oh, D.H., and A.J. Patton. 2017. "Modeling dependence in high dimensions with factor copulas." *Journal of Business & Economic Statistics* 35:139–154.

—. 2016. "Time-varying systemic risk: evidence from a dynamic copula model of CDS spreads." *Journal of Business & Economic Statistics*, pp. 1–47.

Patton, A.J. 2013. "Copula methods for forecasting multivariate time series." In G. Elliott and A. Timmermann, eds. *Handbook of Economic Forecasting, Volume 2*. Oxford: Elsevier.

Pesaran, H.H., and Y. Shin. 1998. "Generalized impulse response analysis in linear multivariate models." *Economics Letters* 58:17–29.

Philipson, T.J., and R.A. Posner. 1999. "The long-run growth in obesity as a function of technological change." Working paper, National Bureau of Economic Research.

Potts, M. 2003. "Overweight and obesity in China." *Policy* 29:163–96.

Price, R.A., D.R. Reed, and N.J. Guido. 2000. "Resemblance for body mass index in families of obese African American and European American women." *Obesity Research* 8:360–366.

Sacerdote, B. 2004. "What happens when we randomly assign children to families?" Unpublished, Working Paper, National Bureau of Economic Research.

Saghaian, S.H. 2010. "The impact of the oil sector on commodity prices: correlation or causation?" *Journal of Agricultural and Applied Economics* 42:477–485.

Swanson, N.R., and C.W. Granger. 1997. "Impulse response functions based on a causal approach to residual orthogonalization in vector autoregressions." *Journal* *of the American Statistical Association* 92:357–367.

Tang, K., and W. Xiong. 2012. "Index investment and the financialization of commodities." *Financial Analysts Journal* 68:54–74.

Taveras, E.M., S.L. Rifas-Shiman, C.S. Berkey, H.R. Rockett, A.E. Field, A.L. Frazier, G.A. Colditz, and M.W. Gillman. 2005. "Family dinner and adolescent overweight." *Obesity Research* 13:900–906.

Vogler, G.P., T. Sørensen, A.J. Stunkard, M. Srinivasan, and D. Rao. 1995. "Influences of genes and shared family environment on adult body mass index assessed in an adoption study by a comprehensive path model." *International Journal of Obesity* 19:40–45.

Wilson, S.E. 2002. "The health capital of families: an investigation of the inter-spousal correlation in health status." *Social Science & Medicine* 55:1157–1172.

Wooldridge, J.M. 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT press.

APPENDIX A

APPENDIX MATERIAL FOR SECTION 2

A.1 The skewed t distribution

As discussed in section 2.2.3, standardization of copula shocks is required in the estimation of dynamic conditional correlation. In this part we introduce the basics of skewed t copula which are used to drive dynamic correlations in our empirical analysis. Demarta and McNeil (2005) show that the skewed t distribution has the following stochastic representation:

$$X = \sqrt{W}Z + \lambda W \tag{A.1}$$

where w is an inverse gamma variable such that $W \sim IG(v/2, v/2)$, Z is a normal variable such that $Z \sim N(0_N, \Gamma)$, and λ is the asymmetry parameter. Equation (A.1) suggests skewed t distribution has a normal mixture structure which implies the expected value of X is:

$$E(X) = E(E[X|W]) = E(W)\lambda = \frac{v}{v-2}\lambda \tag{A.2}$$

and the covariance matrix is:

$$\begin{aligned} Cov(X) &= E(Var(X|W)) + Var(E(X|W)) \\ &= \frac{v}{v-2}\Gamma + \frac{2v^2\lambda^2}{(v-2)^2(v-4)} \end{aligned}$$

The expected value and covariance matrix show how the skewed t distribution is linked with the copula correlation Γ. Standardization of student copula shocks can be implemented using these moments with $\lambda = 0$.

A.2 Out-of-sample model comparison

Out-of-sample period spans from June 1st 2007 to Dec 1st 2009 and May 1st 2014 to May 23rd 2015, which includes 900 days as the test set. We use a fixed rolling window of 1000 days to fit the dynamic copulas and forecast one-day ahead copula density. As the computational burden is still high even we employ composite likelihood in the estimation of dynamic copulas, we estimate each copula model seven times every six months in the test set. These dates are June 1st 2007, Dec 4th 2007, June 3rd 2008, Dec 2nd 2008, June 2nd 2009, May 1st 2014 and Nov 4th 2014. We reestimate the dynamic copula model using 1000 days before each date to predict copula density for the next six months*. To precisely forecast one-day ahead volatility and recover copula shocks, we reestimate every univariate ARIMA-GARCH model for every day in the test test. Suppose period t is one of the seven dates mentioned above, the detailed procedure of OOS model comparison is summarized in the following four steps:

Step 1: estimate univariate ARIMA-GARCH model for each commodity returns and dynamic copulas using data in period [t-1001, t-1] as the training set and save all estimates.

Step 2: predict one-day ahead volatility $\sigma_{i,t}$ and recover the error $\epsilon_{i,t}$ using return $y_{i,t}$ for each i. Yield the cumulative distribution $\eta_{i,t}$ and save the copula density from the estimated dynamic copula.

Step 3: at the end of date t reestimate the univariate ARIMA-GARCH model using data in [t-1000, t] period, and repeat step 2. Note that the copula model is not reesitmated in this step.

*For recent development of sample split selection for comparing forecast ability of time series models, we refer interested readers to Hansen and Timmermann (2012) who propose a test statistic that is robust to the effect of considering multiple sample split points.

Step 4: repeat step 2 and 3 until we approach the next copula reestimation date on which we begin from step 1.